Computer Vision and Image Understanding
journal homepage: www.elsevier.com

Occlusion aware particle filter tracker to handle complex and persistent occlusions (Supplementary Material)

Kourosh Meshgi[a,**], Shin-ichi Maeda[a], Shigeyuki Oba[a], Henrik Skibbe[a], Yu-zhe Li[a], Shin Ishii[a]

[a] *Graduate School of Informatics, Kyoto University, Yoshidahonmachi, Sakyo Ward, Kyoto, 606–8501 JAPAN*

ABSTRACT

This document contains the supplementary material for the paper submitted to “Pattern Recognition Letters”. Qualitative and quantitative analysis of the tracker performance for different scenarios is presented in this document. Additionally the details of the robust feature calculation section is presented in this document.

1. Detailed Analysis of Videos

In this section, the results of all trackers over all five videos used in the paper are presented. Following a brief overview of the dataset configuration and rival algorithms, the videos are presented one-by-one with a quick review on their characteristics. Then the result of the trackers are presented along with a discussion over the performance of them.

As mentioned before, this study uses five videos from Princeton Tracking Dataset (Song and Xiao, 2013). Authors of this paper aimed to standardize a uniform evaluation criteria for tracker comparison, having occlusion evaluation in mind. The dataset contains 100 sequences, acquired with Microsoft Kinect, with 640 × 480 pixels 8-bit RGB color images and same size 8-bit normalized depth map, and tried to cover various types of occlusion with real-world indoor setup (office, room, library, shop, sports, concourse, fields). The targets are comprised of humans, animals or relatively rigid objects and their initial position is provided for the first frame. This dataset is publicly available at: `http://tracking.cs.princeton.edu/`. The five manually annotated videos is employed in this paper are: `new_ex_occ4`, `face_occ_5`, `bear_front`, `child_no1` and `zcup_move_1`. Based on the definition of video challenges presented in Table 2 of a recent benchmark over online object trackers (Wu et al., 2013), we attributed each video as follows:

IV illumination variation;
SV scale variation - when the target grows or shrinks by the factor of 2 comparing to the initial size;
DEF non-rigid deformation or articulation;
MB motion blur;
FM fast motion - when the target displaces more than 20 pixels between two consecutive frames;
IPR in-plane rotation;
OPR out-of-plane rotation;
OV out-of-view;
BC background clutter - the background near the target has similar color or texture;
LR low resolution - target bounding box has less than 400 pixels.

Inspired by a pepaer by Vezzani et al. (2011), the occlusions of the video are divided into following categories:

PTO partial occlusion;
SAO self- or articulation occlusion;
TFO temporal full occlusion - shorter than 3 frames;
PFO persistent full occlusion;
CPO complex partial occlusion - including “split and merge” and permanent changes in a key attribute of a part of target;
CFO complex full occlusion.

As mentioned in the main paper, we compare OAPFT, our proposed method using different feature sets with OI+SVM (Song and Xiao, 2013), STRUCK (Hare et al., 2011) and ACPF (Nummiaro et al., 2003). More information about three former trackers is available in their dedicated webpages `http://ishiilab.jp/member/meshgi-k/oapft.html`, `http://`

[**] Corresponding author: Tel.: +81-75-753-4809;
e-mail: `meshgi-k@sys.i.kyoto-u.ac.jp` (Kourosh Meshgi)
URL: `http://ishiilab.jp/member/meshgi-k/oapft.html` (Kourosh Meshgi)

tracking.cs.princeton.edu/code.html, and http://www.samhare.net/research/struck.

1.1. Sequence 1: new_ex_occ4

Properties

This sequence contains 51 frames, in which a pedestrian walks along a corridor in which she is occluded by another pedestrian for several frames. The provided ground truth of this file is not homogeneous because in some frames it shrinks suddenly to cover the small unoccluded part of the target. A ground truth file is called homogeneous if the size of the target is kept consistent during the tracking and do not change drastically between consecutive frames. Furthermore, the box scales are erroneous and in some cases stretches beyond the size of the image (the corrected version is utilized in this experiment). Additionally the initial bounding box is different from the one mentioned in the ground truth file, thus the values from ground truth is used for the initialization of the trackers throughout this experiment – the case holds for the other videos. This is the most challenging video of these five because of the complex full occlusion in which the target and the occluder are similar in color patterns and shape (and as a result similar HOG), and cross each other so close that the noisy depth values almost have similar values. The sequence can be attributed by DEF, MB, FM and BC.

Analysis of Trackers

The implementation of *ACPF*, the particle filter tracker, although is similar to our *C* tracker, but it also involves geometric weighting of the pixels in the color histogram, that works well in some scenarios, but suffers severely in the presence of background clutter. Another difference is the use of adaptive binning in our implementation of color histogram. We used k-means clustering of input image in first frame with 40 clusters and used the centroid centers as histogram bins in color histogram, where *ACPF* used $8 \times 8 \times 8$ bins for RGB channels which leads to many zero entries in the resulting feature vector. Furthermore, KL-divergance performs better for discriminating color histograms and guides the particle filter far better than Bhattacharyya distance used in *ACPF*.

By having a close look on Figure 1 it is obvious that in Frame 30 with the emergent occlusion and sudden shrinkage of target bounding box, both our proposed method, *OAPFT*, and *OI+SVM* transit into occlusion mode. The occlusion indicator of *OI+SVM* detects a sudden drop in the size of bins around the target nominal depth and declares occlusion. Looking into *OAPFT* internal dynamics reveals that many particles couldn't find a probable candidate for following and occlusion mark particles dominates the population, resulting in occlusion declaration for the tracker. A few frames before occlusion, in frame 27, some of trackers based on the corresponding features still respond to the object or background. For instance tracker *E* (edge-only) converged to a non-target part of the scene due to background edge clutter, which is solved later with more particles going to occlusion state in the succeeding frames.

With the reappearance of the target in frame 32, the occlusion state is resolved because the number of occluded particles and their probability don't exceed the fixed threshold, δ_{occ} and some of the trackers are evoked. Due to the expanded search zone, the estimation of the target is generated from a wide area, where some of them maybe not relevant in the case of cluttered features (e.g. edge) that cause the immediate estimate of the tracker to be less accurate (r.t. trackers *E* and *S* in the frame 32). Tracker *OI+SVM* could not recover from this complex occlusion and maintain the occlusion state for most of the following frames.

If the feature is not expressive enough for the scenario, their corresponding tracker would lose the target, as it is seen for single-featured trackers or a combination of them (e.g. trackers *C*, *E*, *S*, *CE* and *CG* in frame 42). Anyway, the quick recovery from occlusion for well described trackers is evident from frame 40 where trackers start to follow the target again. As mentioned earlier, having enough descriptive power in the form of features, is an essential factor for a successful tracking. For example in this sequence, color, edges and shape features and the pure combinations are not adequate to describe the target and the corresponding tracker chase the background or similar object. But when features with covering different aspects of the object are teamed up with each other, the weakness of each is covered e.g. *CGS*. Also there is a high probability that one of the features can solve the tracking video almost well, and in combination with other features, its accuracy could be improved. This is the case observed for depth feature in this videos. Finally it should be mentioned that including more features is not always a good solution as it might weaken the positive effect of successful features in the video. For instance tracker *CDEST* becomes unstable and is unable to find the target quickly after a partial occlusion. This emphasizes the fact that feature selection still has many depths to dive in. Table 1 compares all the algorithms and provide more insight in this regards by comparing the *AUC* and *CPE* values.

1.2. Sequence 2: face_occ_5

Properties

This sequence involves 330 frames, having human face as the target, and a book is moved in front of the target to cover it partially from different angles as well as causing a full occlusion. The background is an office with moving people, but the target remains still throughout the video. The provided ground truth of this video is not homogeneous because its size changes to accommodate the partial occlusions and tries to embody only the visible portion of the target. This sequence contains persistent occlusion (PFO), and several partial occlusions (PTO) and the background has clutter and other distractions (BC).

Analysis of Trackers

The result of the experiment is illustrated in Figure 2. In this video, the geometric distribution of color pixels used in *ACPF* caused this tracker to deviate from the target as it is evident in frame 39. Additionally due to presence of many weak edges in the background, tracker *E* was not successful in tracking the target as it is observed that the tracker lost the target in frame 39 and although found it later in frame 60, did not track the target well due to the template corruption. In frame 142 with another partial occlusion, *ACPF* and *S* lost the target completely and were not able to recover it again.

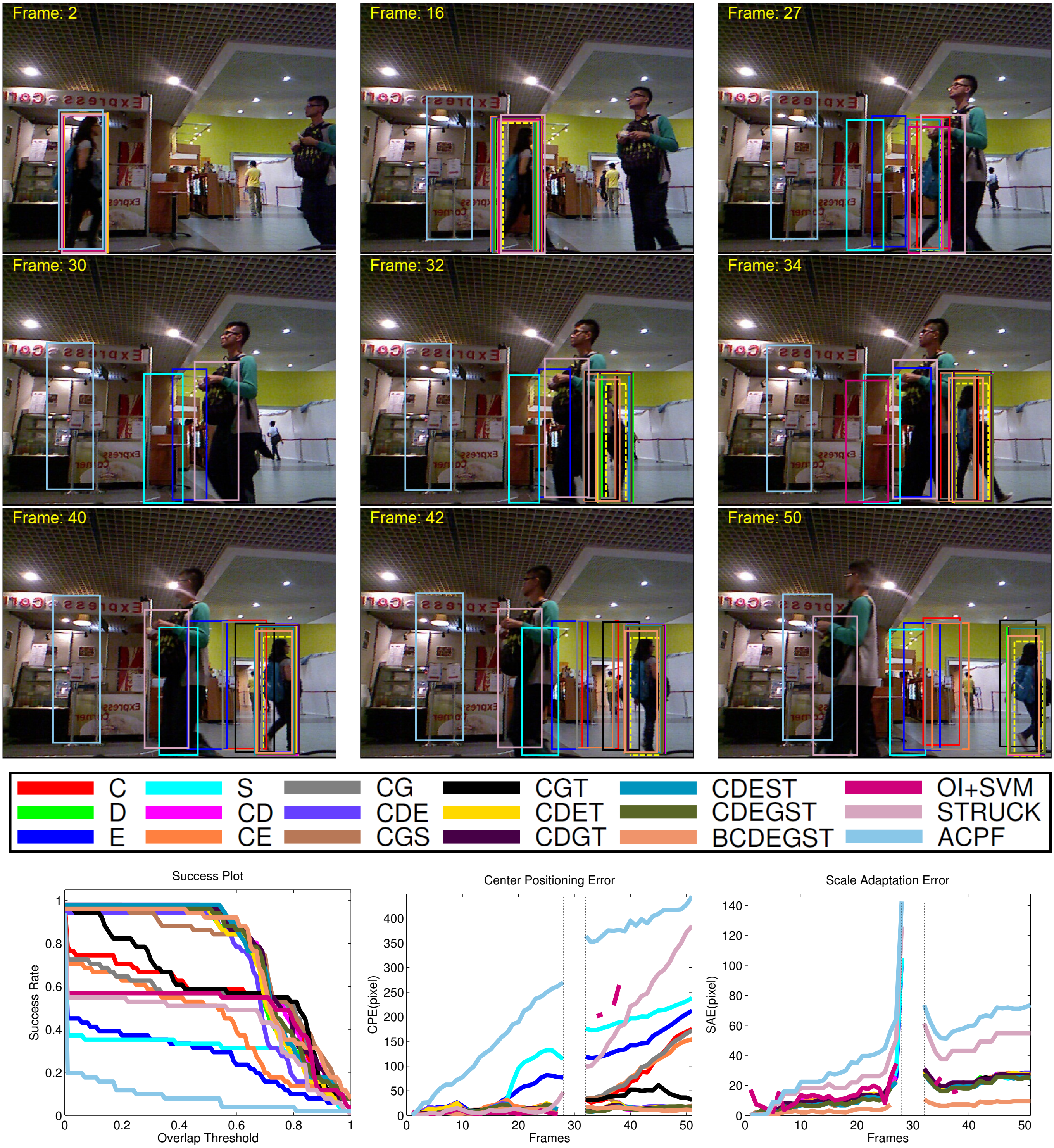

Figure 1. Qualitative Analysis of sequence `new_ex_occ4`, the ground truth is marked with yellow dashed line.

Approaching to full occlusion in frame 166, most of the versions of our proposed method along with *OI+SVM* start to transit to occlusion state. In frame 173, in the middle of persistent occlusion however, it is observed that trackers *S*, *E*, *CE*, *CDE*, *CDEST*, and *ACPF* were not in occlusion state. Looking at the dynamics of these trackers, it is evident that edges and 3D shape fails to describe the target clearly and histogram of color is mainly attracted to the color of skin and doesn't provide high quality description of the target. This is why these trackers fails to track the target and the fixed l_{occ} used for all videos, was not completely suitable for this video. However, with more descriptive features added to each of them, the tracker works well and detect occlusion successfully (e.g. *CDE* → *CDET* and *CDEST* → *CDEGST*).

Table 1. Tracker evaluation for sequence `new_ex_occ4`.

Tracker	cc[#]	*AUC*	*CPE*	*SAE*	*MI*	*FT*	*MT*	*FPS*
C	🟥	55.71	39.75	14.43	**0.0**	1.0	10.0	7.1
D	🟩	71.29	11.09	13.67	**0.0**	3.0	**0.0**	13.4
E	🟦	27.20	78.36	16.16	3.0	**0.0**	25.0	4.4
S	🟦	31.20	103.36	15.17	3.0	**0.0**	29.0	0.6
CD	🟪	**76.00**	10.63	14.81	**0.0**	1.0	**0.0**	6.7
CE	🟧	42.98	42.80	13.45	**0.0**	3.0	12.0	3.2
CG	⬜	53.27	41.21	13.91	**0.0**	2.0	12.0	4.8
CDE	🟪	68.41	13.21	13.34	**0.0**	3.0	**0.0**	3.0
CGS	🟫	74.67	12.16	12.89	**0.0**	1.0	**0.0**	0.6
CGT	⬛	60.90	20.77	13.94	**0.0**	2.0	**0.0**	2.6
CDET	🟨	71.57	12.14	13.78	**0.0**	2.0	**0.0**	2.0
CDGT	🟪	75.22	10.51	14.10	**0.0**	2.0	**0.0**	2.6
CDEST	🟦	74.71	10.83	12.70	**0.0**	1.0	**0.0**	0.5
CDEGST	🟩	73.92	10.72	12.17	**0.0**	2.0	**0.0**	0.5
BCDEGST	🟧	74.57	**8.71**	**4.55**	**0.0**	2.0	**0.0**	0.5
OI+SVM	🟥	48.76	20.28	8.29	1.0	17.0	4.0	<0.1
STRUCK	🟪	42.92	96.66	31.37	3.0	**0.0**	20.0	**17.2**
ACPF	🟦	8.73	228.31	141.57	3.0	**0.0**	38.0	0.9

[#] *cc – Color Code for the Tracker*

Table 2. Tracker evaluation for sequence `face_occ_5`.

Tracker	cc[#]	*AUC*	*CPE*	*SAE*	*MI*	*FT*	*MT*	*FPS*
C	🟥	42.37	43.58	5.46	**0.0**	1.0	6.0	9.7
D	🟩	76.23	8.62	9.26	**0.0**	2.0	**0.0**	**14.6**
E	🟦	18.35	133.76	9.32	14.0	**0.0**	148.0	10.1
S	🟦	25.48	139.25	11.16	14.0	**0.0**	148.0	1.7
CD	🟪	67.90	12.26	9.25	**0.0**	2.0	**0.0**	11.3
CE	🟧	29.20	107.65	9.45	14.0	**0.0**	148.0	7.2
CG	⬜	31.75	70.97	9.19	**0.0**	3.0	120.0	7.8
CDE	🟪	65.38	11.93	9.60	14.0	**0.0**	**0.0**	7.8
CGS	🟫	42.94	**4.27**	**4.10**	**0.0**	149.0	**0.0**	1.5
CGT	⬛	41.34	36.61	9.19	**0.0**	3.0	**0.0**	5.0
CDET	🟨	69.45	11.81	9.11	**0.0**	3.0	**0.0**	5.1
CDGT	🟪	69.25	11.59	9.39	**0.0**	1.0	**0.0**	5.1
CDEST	🟦	39.20	78.77	10.97	14.0	**0.0**	148.0	1.4
CDEGST	🟩	74.19	9.05	10.72	**0.0**	3.0	**0.0**	1.3
BCDEGST	🟧	**79.16**	6.30	10.85	**0.0**	3.0	**0.0**	1.3
OI+SVM	🟥	66.90	5.00	8.94	**0.0**	57.0	**0.0**	<0.1
STRUCK	🟪	41.55	81.33	9.33	14.0	**0.0**	148.0	11.3
ACPF	🟦	28.93	55.53	25.84	14.0	**0.0**	40.0	1.9

[#] *cc – Color Code for the Tracker*

With the target reappearance in frame 183, the occlusion recovery process starts. As it is illustrated in frame 184 most of the trackers that detected occlusion earlier, recovered from the occlusion. However the histogram of color suffered from corrupted template thus trackers *C* and *CG* were unable to track the target successfully afterwards. Moreover due to the same problem, tracker *CGT* fails later in frame 280. Note that tracker *CGS* although detects the occlusion correctly, but stays in occlusion status till the end of scenario.

Table 2 supports the qualitative observations. By a close look at *CPE* and *MT* values, it is evident that trackers with high errors lost the target during the tracking. Also tracker *CGS* with a large *FT* value maintains the occlusion status for many frames during the video. In this video, *OI+SVM* suffers from false tracking state and *STRUCK* lost the target during the persistent full occlusion.

1.3. Sequence 3: `bear_front`

Properties

This sequence has 281 frames, in which a teddy bear is being moved around the screen over a same color background. During the course of this sequence a white box is also moved in a random to cover a part or whole body of the bear several times. The provided ground truth of this file is not homogeneous because the ground truth box shrinks size in the case of partial occlusion. This video contains 6 full occlusions from the same occluder with various lengths and directions. The sequence can be attributed by IV, IPR and BC.

Analysis of Trackers

Figure 3 depicts a few snapshots of the tracker performances handling occlusions. In frame 43, during the full occlusion it is observed that trackers *E* and *CE* fails to detect the occlusion. This failure is rooted in failure of edge detector to find strong persistent edges throughout the sequence, also due to background color clutter, the color feature is not strong enough to prevent tracker *CE* from losing the target. This argument hols for *ACPF* that solely relies on color cues to track the target and get absorbed to the background in this scenario. Additionally *STRUCK* lost the target permanently during this occlusion and tracks the occluder afterwards as it is seen in frame 58. Due to corrupted template, trackers *C* and *CE* in frame 58 is seen to stick to the background, and fail to recover the target.

Furthermore the trackers *CGS* and *CDE* faced problems in tracking as it is seen in frame 68, while the former recovers the target slowly due to lack of enough expressive power of its features, the latter has a corrupted template. Moreover tracker *S* fails to recover from the occlusion completely as shown in frame 220 and lose the target later in this scenario.

Investigating evaluation Table 3, the above mentioned observations are proved as tracker *E* has a high *CPE* and *MT* values, *CE* almost has the same situation, late recovery of *CDE* impacts its *SAE* and *CPE* vlues and *CGS* has a high *MT* value. In this scenario *OI+SVM* performs well thanks to its object detector scheme.

1.4. Sequence 4: `child_no1`

Properties

This sequence has 164 frames, in which the subject moves across a furnished room and crouch in the end. The provided

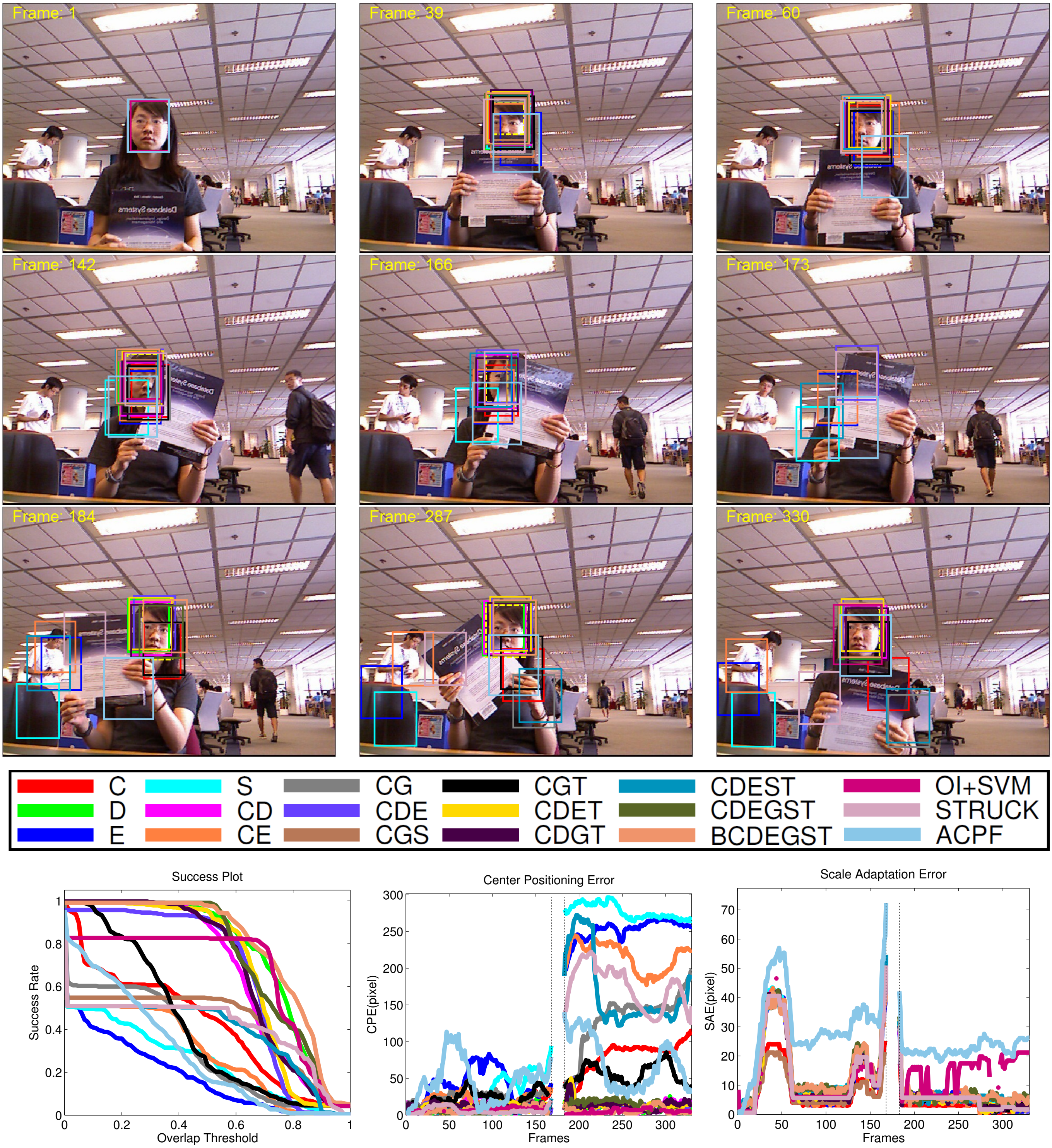

Figure 2. Qualitative Analysis of sequence `face_occ_5`, the ground truth is marked with yellow dashed line.

ground truth of this file is homogeneous. This video involves articulated body motions and deformations, and self-occlusions of the target. While the target is easily distinguishable using color features, other features such as depth suffers from heavy noise and clutter. The sequence can be attributed as by DEF, OPR, and BC.

Analysis of Trackers

As it is seen in Figure 4, the trackers generally perform well over this sequence. This sequence has a depth clutter with many outlier values and noise in the measurements and same depth objects in the field of view, hence tracker *D* with only histogram of depth as the leading cue lost the target early in the scenario. Additionally the background contains various

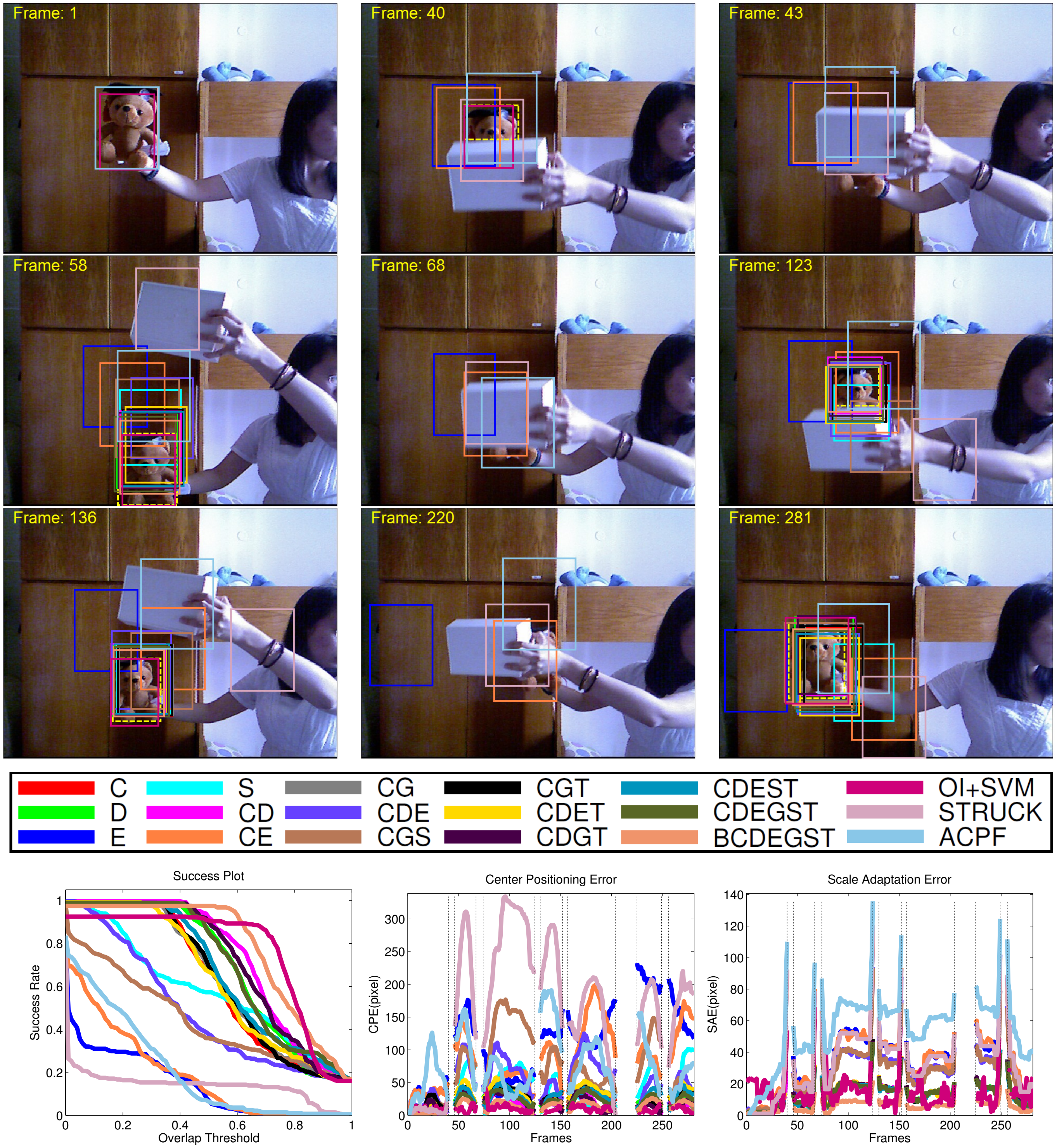

Figure 3. Qualitative Analysis of sequence `bear_front`, the ground truth is marked with yellow dashed line.

complicated structures and details which introduces many unwanted elements into edge and gradient feature-space that distract trackers using such kind of cues. For instance tracker *E* and *CGS* suffers from this clutter and lost the target in early frames on the sequence. As an advanced feature relying on depth information, 3D shape also fails to provide robust information to tracker thus tracker *S* was not successful to track the targets. Tracker *CDEST* with three ineffective features, struggles to track the target but the poor performance of this tracker is observed throughout the sequence. Although the rival algorithms track the target successfully, they don't maintain an optimal size for their target boxes. *OI+SVM* and *ACPF* without explicit size adaptation mechanism and *STRUCK* with no such mechanism, are unable to the scale change of the target dur-

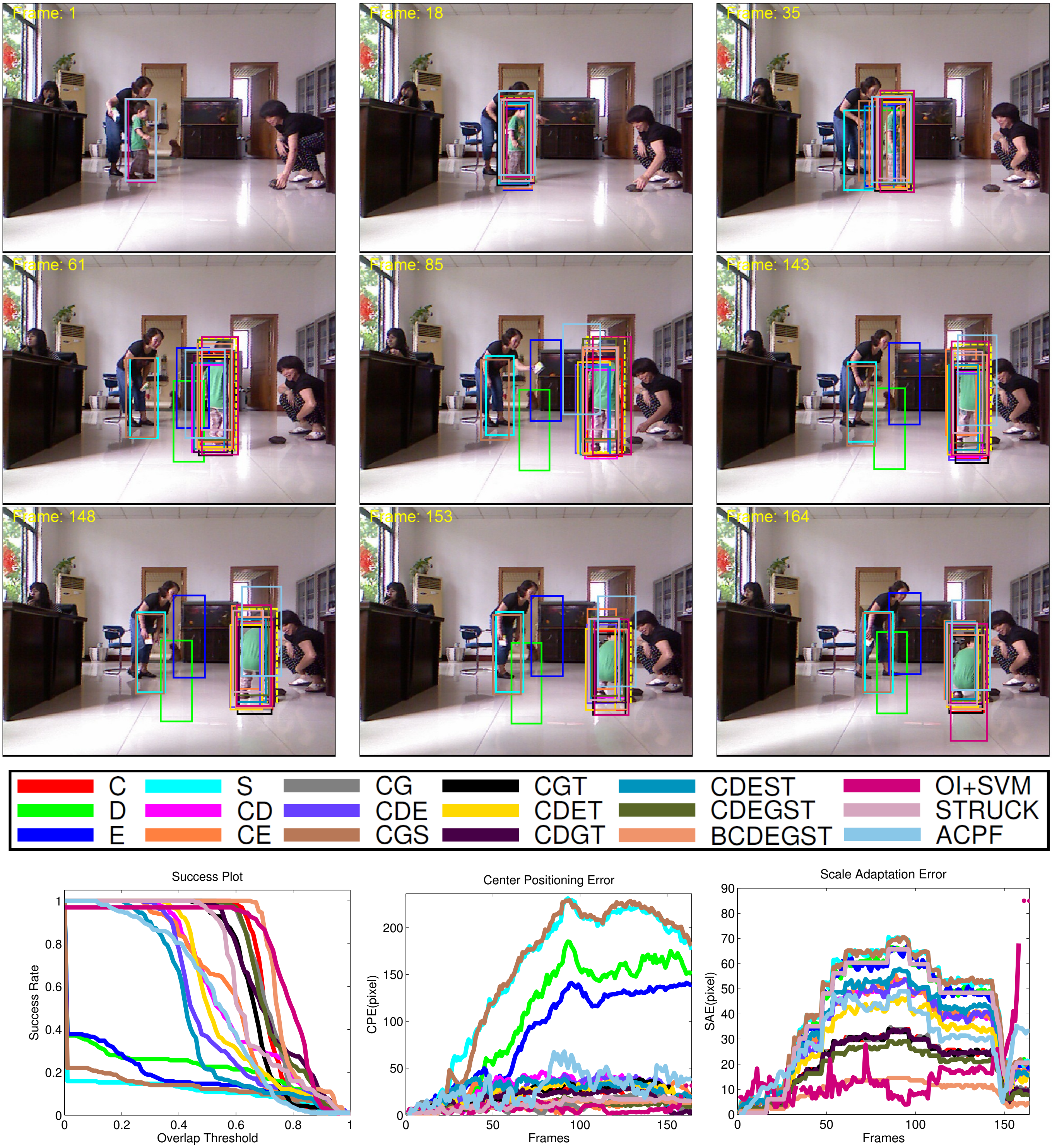

Figure 4. Qualitative Analysis of sequence `child_no1`, the ground truth is marked with yellow dashed line.

ing the scenario. This is especially evident in frame 143. This can be inferred from Table 4 easily. On the other hand tracker *BCDEGST* benefits from "2D Projection Confidence" feature and performs well regarding scale adaptation.

1.5. Sequence 5: zcup_move_1

Properties

This sequence has 370 frames, in which a pot is moved in different directions first and then away from the camera. The camera in this scenario is moving, and the lighting condition is poor. The provided ground truth of this file is homogeneous. Moving camera, same color background and out-of-plane ro-

Table 3. Tracker evaluation for sequence `bear_front`.

Tracker	cc#	*AUC*	*CPE*	*SAE*	*MI*	*FT*	*MT*	*FPS*
C	🟥	63.85	22.81	10.84	**0.0**	2.0	**0.0**	6.1
D	🟩	70.02	18.43	10.78	**0.0**	3.0	**0.0**	**13.4**
E	🟦	14.58	90.48	28.48	46.0	**0.0**	94.0	7.8
S	🟦	58.55	33.13	10.68	**0.0**	3.0	**0.0**	0.8
CD	🟪	74.14	14.98	10.95	**0.0**	1.0	**0.0**	6.0
CE	🟧	19.20	72.98	29.02	46.0	**0.0**	37.0	4.3
CG	⬜	64.56	22.28	10.78	**0.0**	3.0	**0.0**	4.6
CDE	🟪	48.70	45.47	22.28	**0.0**	1.0	**0.0**	4.2
CGS	🟫	44.84	61.16	21.80	**0.0**	3.0	33.0	0.7
CGT	⬛	64.39	22.11	10.70	**0.0**	3.0	**0.0**	2.7
CDET	🟨	64.13	24.54	10.85	**0.0**	2.0	**0.0**	2.7
CDGT	🟪	72.53	15.46	10.76	**0.0**	3.0	**0.0**	2.7
CDEST	🟦	67.40	21.10	10.65	**0.0**	3.0	**0.0**	0.7
CDEGST	🟩	70.85	17.63	10.75	**0.0**	3.0	**0.0**	0.6
BCDEGST	🟧	**78.90**	10.86	**4.76**	**0.0**	7.0	**0.0**	0.6
OI+SVM	🟥	76.99	**7.84**	11.14	1.0	20.0	**0.0**	2.1
STRUCK	🟪	14.12	142.03	28.07	46.0	**0.0**	154.0	9.4
ACPF	🟦	21.01	66.43	45.52	46.0	**0.0**	23.0	0.9

cc – Color Code for the Tracker

Table 4. Tracker evaluation for sequence `child_no1`.

Tracker	cc#	*AUC*	*CPE*	*SAE*	*MI*	*FT*	*MT*	*FPS*
C	🟥	72.12	18.36	20.07	**0.0**	**0.0**	**0.0**	9.5
D	🟩	22.22	104.67	39.64	**0.0**	**0.0**	103.0	**16.1**
E	🟦	17.72	86.55	39.57	**0.0**	**0.0**	101.0	9.0
S	🟦	11.58	150.53	43.53	**0.0**	**0.0**	138.0	1.6
CD	🟪	59.59	26.68	32.11	**0.0**	**0.0**	**0.0**	9.5
CE	🟧	59.05	15.61	32.68	**0.0**	**0.0**	**0.0**	6.4
CG	⬜	71.56	12.09	19.91	**0.0**	**0.0**	**0.0**	7.1
CDE	🟪	52.48	26.75	32.15	**0.0**	**0.0**	**0.0**	6.6
CGS	🟫	13.20	151.24	43.40	**0.0**	**0.0**	128.0	1.5
CGT	⬛	67.12	22.73	19.91	**0.0**	**0.0**	**0.0**	4.5
CDET	🟨	56.84	22.52	28.09	**0.0**	**0.0**	**0.0**	4.4
CDGT	🟪	72.04	16.98	19.95	**0.0**	**0.0**	**0.0**	4.4
CDEST	🟦	46.90	27.13	34.98	**0.0**	**0.0**	**0.0**	1.2
CDEGST	🟩	73.16	10.52	17.77	**0.0**	**0.0**	**0.0**	1.2
BCDEGST	🟧	77.21	10.92	**9.15**	**0.0**	**0.0**	**0.0**	1.2
OI+SVM	🟥	**77.30**	**6.81**	13.97	**0.0**	5.0	**0.0**	<0.1
STRUCK	🟪	66.41	11.73	39.55	**0.0**	**0.0**	**0.0**	11.8
ACPF	🟦	53.83	29.86	30.31	**0.0**	**0.0**	**0.0**	1.6

cc – Color Code for the Tracker

tation which causes self-occlusions are the most prominent attributes of this sequence (OPR, BC).

Analysis of Trackers

A glance at Figure 5 reveals that the target is hard to distinguish from the background only using color feature. Due to this reason *ACPF* tracker fails to track the object from the early frames of the video (i.e. frame 19). Although tracker *C* benefits from color histogram with adaptive binning, but it also suffers from this problem and is unable to keep the main focus on the target (r.t frames 209, 242, 257). Additionally due to hard edges introduced by the background and soft edges of the target, trackers *E* and *CE* show poor tracking performance during the video. Additionally tracker *CGT* performs not very well for this sequence (e.g. frame 257).

Evaluation Table 5 supports the claim that color, edge and 3D shape features fails in this scenario as it is inferred from trackers *C*, *E* and *S* respectively. On the other hand histogram of depth works very well for the scenarios (r.t tracker *D*) and is effectively improved with the combination of other features in trackers *CD*, *CDE* and *CDET*. Also the HOG feature was effective not only in localizing the target, but also in enforcing the scale adaptation to the system as it is observed in trackers *CG* and *CDGT*. However the poor performance of 3D shape tracker enforce tracker *CGS* to stop tracking (indicated by high *FT* value) as the probability of all non-occluded particles become very small. Moreover this feature reduce the performance of the trackers including it as it is seen from tracker *CDEST* against *CDET*.

The most important observation in this table is the failure of 2D confidence projection feature to improve the scale adaptation of tracker *BCDEGST*. This is due to the fact that the background subtraction employed here (inspired from Lo and Velastin (2001)), assumes a static background, the condition which is is violated by the camera movement in this scenario. However using the proposed feature normalization procedure, the impact of this failure is significantly reduced on the tracker scale adaptation.

Detector based tracker such as *STRUCK* and *OI+SVM* performs well in this scenario, while the color-based particle filter tracker, *ACPF*, suffer from low contrast. This low contrast cause non-adaptive bin color histograms to experience value concentration on a few number of bins that make it difficult to separate resembling bounding box from irrelevant ones.

2. Video Demonstration

In order to better illustrate the comparison between trackers, supplementary videos are available in the first author webpage: `http://ishiilab.jp/member/meshgi-k/oapft.html`. These videos compare the performance of different feature sets for our proposed tracker, visualization of different features extracted from sequence, and comparison between our algorithm and *ACPF*, *OI+SVM* and *STRUCK*.

3. Robust Feature-based Template Matching

A feature-based template matching with unnormalized features is prone to following errors:

- Feature Noise: the calculated values of feature contains noise due to noise in raw input (e.g. depth range data) or during the calculation.

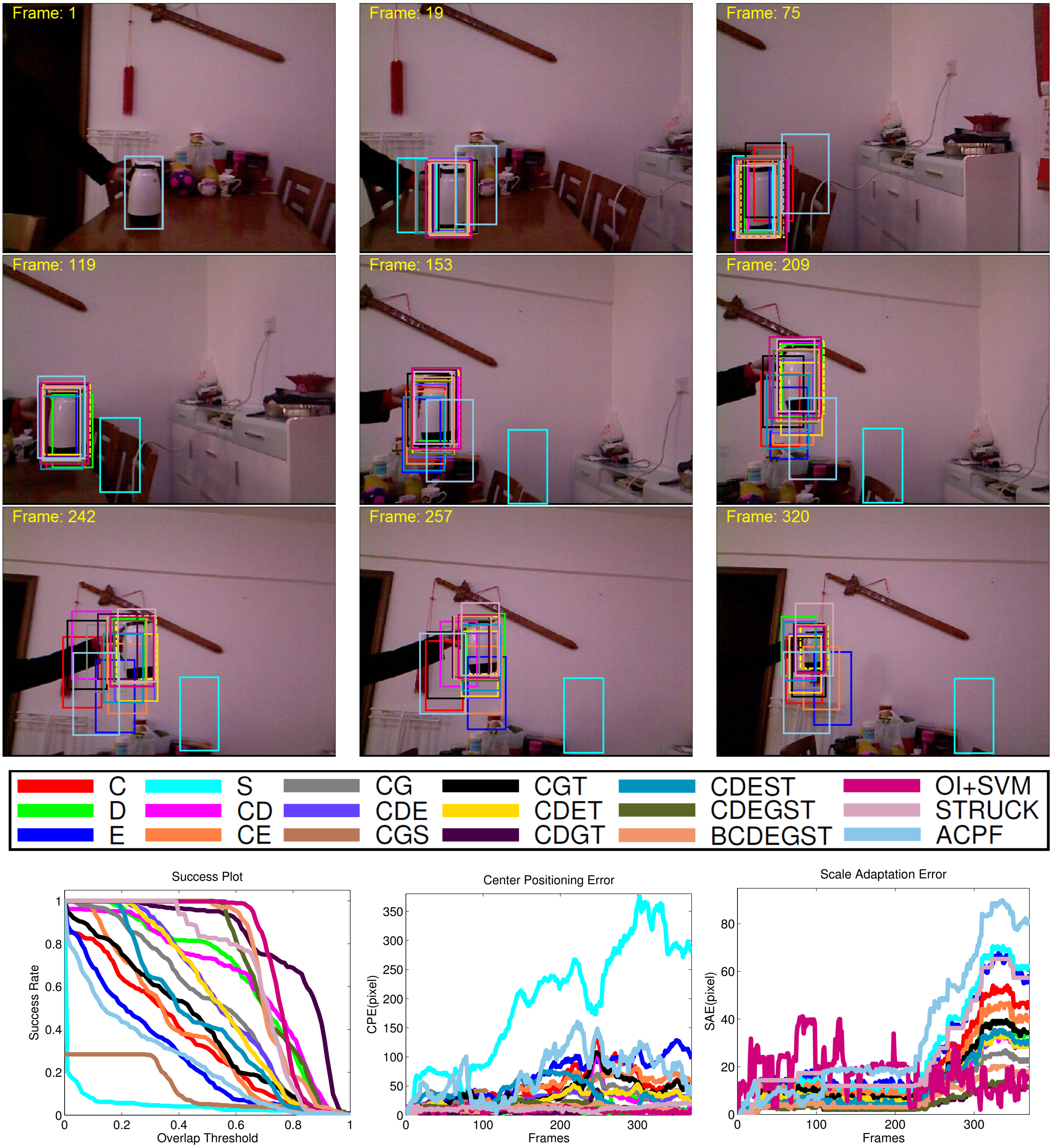

Figure 5. Qualitative Analysis of sequence `zcup_move_1`, the ground truth is marked with yellow dashed line.

- Zero Likelihood: if the calculated value of a feature is very far from the template, the likelihood of this feature calculated from the negative exponential likelihood function becomes zero. According to equation (4) of the main paper, this zero value results to a zero likelihood for the particle, no matter how the other features are measuring the distance.

- Feature Failure: if the distance of all particles from template are very large, for example due to illumination change, all of the likelihood will be zero that impedes tracker completely. This case usually happens when an essential assumption under which a feature is extracted is violated.

- Domination/Value Range Mismatch: if the resulting dis-

Table 5. Tracker evaluation for sequence `zcup_move_1`.

Tracker	cc[#]	*AUC*	*CPE*	*SAE*	*MI*	*FT*	*MT*	*FPS*
C	🟥	35.89	47.70	19.42	**0.0**	**0.0**	41.0	8.4
D	🟩	65.35	14.84	14.50	**0.0**	**0.0**	**0.0**	14.4
E	🟦	27.59	65.61	24.61	**0.0**	**0.0**	40.0	9.9
S	🟦	5.22	194.09	24.70	**0.0**	**0.0**	291.0	1.4
CD	🟪	64.96	16.75	12.04	**0.0**	**0.0**	12.0	8.0
CE	🟧	38.62	44.26	17.24	**0.0**	**0.0**	**0.0**	6.7
CG	⬜	52.68	26.49	9.79	**0.0**	**0.0**	**0.0**	6.2
CDE	🟪	56.04	25.78	12.29	**0.0**	**0.0**	**0.0**	6.7
CGS	🟫	12.82	8.96	**1.39**	**0.0**	265.0	**0.0**	1.3
CGT	⬛	41.93	38.08	14.54	**0.0**	**0.0**	19.0	4.0
CDET	🟨	55.40	26.41	12.24	**0.0**	**0.0**	**0.0**	4.2
CDGT	🟪	**81.67**	**6.95**	4.85	**0.0**	**0.0**	**0.0**	4.0
CDEST	🟦	46.50	33.20	11.93	**0.0**	**0.0**	**0.0**	1.1
CDEGST	🟩	72.57	12.21	4.86	**0.0**	**0.0**	**0.0**	1.1
BCDEGST	🟧	72.67	11.16	7.27	**0.0**	**0.0**	**0.0**	1.0
OI+SVM	🟪	75.79	8.46	17.88	**0.0**	1.0	**0.0**	<0.1
STRUCK	🟪	68.36	11.96	24.70	**0.0**	**0.0**	**0.0**	**17.3**
ACPF	🟦	25.23	71.76	33.12	**0.0**	**0.0**	54.0	1.6

[#] *cc – Color Code for the Tracker*

tances of different feature are in different scales, the combination of such numbers requires a normalization process.

Following the notation of the main paper, especially equation (4), we assume M features, N particles and the set $\mathcal{J}_t^{nocc}$ are non occluded particles at time t. We also introduce $d_i(Y_{j,t})$ as:

$$d_i(Y_{j,t}) = \frac{D_i(f_i(Y_{j,t}), \theta_{i,t})}{\sigma_i} \;, \; i = 1, \ldots, M \tag{1}$$

In order to solve range mistamtch/domination and normalize feature values, a normalization process can be applied on non-occluded particles independently for each feature:

$$0 \le \frac{k d_i(Y_{j,t})}{\sum_{j \in \mathcal{J}_t^{nocc}} d_i(Y_{j,t})} \le k \tag{2}$$

where k is a constant coefficient (here we chose $k = 10$). This normalization however is sensitive to feature noise and outliers, and the features large distance from template are causing a zero likelihood of the feature, which disables the particle completely. In order to solve this problem we introduced a regularization term to feature normalization, i.e.

$$d_i'(Y_{j,t}) = \frac{d_i(Y_{j,t}) + \eta_i}{\left(\sum_{j \in \mathcal{J}_t^{nocc}} d_i(Y_{j,t})\right) + \eta_i} \tag{3}$$

in which η_i is the regularization term for each feature and is a function of ϵ and particle distances to template. The new $d_i'(Y_{j,t})$ can then be used in equation (4) of the main paper,

$$p\left(Y_{j,t} | B_{j,t}, Z_{j,t} = 0, \theta_t\right) \propto \prod_{i=1}^{M} \exp\left(-d_i'(Y_{j,t})\right),$$

and by normalizing the likelihood values for all the particles in the range of $[\epsilon, 1]$ the probability of each particle is calculated.

Acknowledgments

The authors are grateful for the assistance given by Sam Hare and Amir Saffari for their technical assistance with *STRUCK* tracker.

References

Hare, S., Saffari, A., Torr, P.H., 2011. Struck: Structured output tracking with kernels, in: ICCV, 2011 IEEE Intl. Conf., IEEE. pp. 263–270.

Lo, B., Velastin, S., 2001. Automatic congestion detection system for underground platforms, in: Intelligent Multimedia, Video and Speech Processing, 2001. Intl. Symp., IEEE. pp. 158–161.

Nummiaro, K., Koller-Meier, E., Van Gool, L., 2003. An adaptive color-based particle filter. J. Image and Vision Computing 21, 99–110.

Song, S., Xiao, J., 2013. Tracking revisited using rgbd camera: Unified benchmark and baselines, in: ICCV 2013. IEEE Intl Conf., IEEE.

Vezzani, R., Grana, C., Cucchiara, R., 2011. Probabilistic people tracking with appearance models and occlusion classification: The ad-hoc system. Pattern Recognition Letters 32, 867–877.

Wu, Y., Lim, J., Yang, M.H., 2013. Online object tracking: A benchmark, in: CVPR, 2013 IEEE Conf., IEEE. pp. 2411–2418.