Fast-deepKCF Without Boundary Effect
-Supplementary Material-

Linyu Zheng, Ming Tang, Yingying Chen, Jinqiao Wang, Hanqing Lu
National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
University of Chinese Academy of Sciences, Beijing, China
{linyu.zheng, tangm, yingying.chen, jqwang, luhq}@nlpr.ia.ac.cn

Abstract

In this supplementary material, we will provide some additional experimental results, running speed analysis, and possible future work.

1. Attribute Evaluation on OTB

The sequences in OTB-2015 [13] are annotated with eleven attributes for further analyzing the performance of trackers in different aspects. We select eight representative attributes [1] to further compare our fdKCF* with the two groups of modern correlation filter based trackers (CF trakcers) which are listed in our submitted paper.

Fig. 1 and Table 2 show the comparison of our fdKCF* with the first group of CF trackers which can run at beyond 20fps. On all four criteria and eight representative attributes, our fdKCF* significantly outperforms all other real-time CF trackers by large margins.

Fig. 2 and Table 3 show the comparison of our fdKCF* with the second group of CF trackers which produce state-of-the-art localization accuracy but unnecessarily run at real-time speeds. On the whole, the localization accuracy of our fdKCF* is competitive with that of ECO [2] and remarkably superior to other state-of-the-art CF trackers.

2. Qualitative Evaluation on OTB

Fig. 3 illustrates the tracking results of our fdKCF* and five representative CF trackers, including MKCFup [11], ECO [2], BACF [5], SRDCF [3] and C-COT [4], on six hard and very hard sequences [6] of OTB-2015. Our fdKCF* is superior to other trackers.

3. Running Speed Analysis

We report the detailed running time of each main component in our fdKCF* under the assumption that $H = W = 60$, $h = w = 15$ and $C = 96 + 512$ where the target size and the search region size are 60×60 and 240×240, respectively, the cell sizes of features are 4×4, and the channels of shallow level features and deep features are 96 and 512, respectively. In fact, according to the Implementation Details of submitted paper, this is the assumption of maximum target size and search region size and therefore the case of slowest running speed. All tests are conducted on a single TITAN X GPU. Notably, in the experiments of submitted paper, fdKCF* employs the linear kernel.

Feature Extraction	fCKM	Gauss-Seidel for α^*
4ms	9ms	8ms

(a) Running time of Training.

Feature Extraction	Recover w	Calculate Response
15ms	3ms	8ms

(b) Running time of multi-scale detection with linear kernel.

Table 1: Running time of each component in our fdKCF* with $H = W = 60$, $h = w = 15$, and $C = 608$.

Table 1 reports the results. It is seen from the table that our fCKM consumes less in the total running time. In addition, according to the submitted paper, it is clear that the running time of fCKM is still less (about 12ms) even though the number of feature channels is up to 1024. These tests further confirm the following three facts. (i) Our fCKM is efficient even though the high-dimensional deep features are employed. (ii) The fps of our fdKCF* is almost independent of the number of feature channels because all components whose time-consuming is related to the number of feature channels are efficient and are very little affected by the number of feature channels. (iii) The training speed of our fdKCF* is fast, *i.e*. about 20ms (including features extraction) each frame in the slowest case.

4. Future Work

We present the possible future work to improve our fdKCF* from two aspects: discrimination power and localiza-

[1] deformation, fast motion, in-plane rotation, illumination variation, low resolution, motion blur, out-of-plane rotation, and scale variation

tion accuracy.

4.1. Discrimination Power

It is well known that the classic CF tracker ECO [2] employs many techniques and tricks including adding handcrafted features, sparse updating, and feature dimension reducing to improve the discrimination power and running speed of its baseline tracker C-COT [4]. The efficient tracking framework of our novel fdKCF* is fundamentally different from that of C-COT, and the running speed of our fdKCF* is about 80 times that of C-COT in the case of as fair a comparison as possible. We believe that most of the same ideas above to improve the discrimination power of C-COT in ECO can also be applied to our efficient fdKCF* to improve it further.

In addition to the above, there are a lot of improvements based on KCF [8] to improve its discrimination power in recent years, such as CFWCR [7], MCCT [12], MKCFup [11], and the winner of recent VOT2018 challenges [9], LADCF [14], are also based on KCF. All these trackers employ FFT to accelerate its computation resulting in the negative boundary effect. And there is no one which can run in real-time speed when both relaxing the boundary effect and employing the high-dimensional deep features are executed at the same time. On the contrary, our novel fdKCF* employ the proposed fCKM to accelerate its computation in spatial domain without the boundary effect. And it can run at real-time when the high-dimensional deep features are employed. Therefore, we believe that most of the same ideas above to improve the discrimination power of KCF can also be applied to our efficient fdKCF* to improve it further, and the resulting trackers are probably more efficient and more robust than their original KCF based ones.

4.2. Localization Accuracy

Recently, bounding box regression based methods have been widely used to refine the raw tracking results of base trackers in visual object tracking. Typically, SiamRPN [10], DaSiamRPN [15], and ATOM [1] achieve high localization accuracy on public datasets, although the robustness and the discrimination power of their base trackers are relatively poor compared to other state-of-the-art trackers.

Our fdKCF* does not resort to the bounding box regression based methods to refine its tracking results in our current implementation. Thanks to ATOM [1] which proposes an efficient and general bounding box regression method to visual object tracking, and it can be employed in our fdKCF* to improve its localization accuracy without obstacles. It is worth noting that when ATOM is employed, our fdKCF* does not need to perform multi-scale detection, and its single-scale detection version can run at beyond 60 fps.

References

[1] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Atom: Accurate tracking by overlap maximization. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[2] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, Michael Felsberg, et al. Eco: Efficient convolution operators for tracking. In *CVPR*, volume 1, page 3, 2017.

[3] Martin Danelljan, Gustav Hager, Fahad Shahbaz Khan, and Michael Felsberg. Learning spatially regularized correlation filters for visual tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4310–4318, 2015.

[4] Martin Danelljan, Andreas Robinson, Fahad Shahbaz Khan, and Michael Felsberg. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision*, pages 472–488. Springer, 2016.

[5] Hamed Kiani Galoogahi, Ashton Fagg, and Simon Lucey. Learning background-aware correlation filters for visual tracking. In *ICCV*, pages 1144–1152, 2017.

[6] Bohyung Han, Jack Sim, and Hartwig Adam. Branchout: Regularization for online ensemble tracking with convolutional neural networks. In *Proceedings of IEEE International Conference on Computer Vision*, pages 2217–2224, 2017.

[7] Zhiqun He, Yingruo Fan, Junfei Zhuang, Yuan Dong, and HongLiang Bai. Correlation filters with weighted convolution responses. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1992–2000, 2017.

[8] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):583–596, 2014.

[9] Matej Kristan, Ales Leonardis, Jiri Matas, Michael Felsberg, Roman Pfugfelder, Luka Cehovin Zajc, Tomas Vojir, Goutam Bhat, Alan Lukezic, Abdelrahman Eldesokey, Gustavo Fernandez, and et al. The visual object tracking vot2018 challenge results, 2018.

[10] Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, and Xiaolin Hu. High performance visual tracking with siamese region proposal network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8971–8980, 2018.

[11] Ming Tang, Bin Yu, Fan Zhang, and Jinqiao Wang. High-speed tracking with multi-kernel correlation filters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4874–4883, 2018.

[12] Ning Wang, Wengang Zhou, Qi Tian, Richang Hong, Meng Wang, and Houqiang Li. Multi-cue correlation filters for robust visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4844–4853, 2018.

[13] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1834–1848, 2015.

[14] Tianyang Xu, Zhen-Hua Feng, Xiao-Jun Wu, and Josef Kittler. Learning adaptive discriminative correlation filters via

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.752	0.711	0.735	0.615	0.519	0.643	0.603	0.579
mPN	0.767	0.716	0.701	0.609	0.511	0.670	0.596	0.615

(a) Deformation.

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.834	0.767	0.761	0.622	0.545	0.624	0.703	0.614
mPN	0.798	0.747	0.748	0.607	0.530	0.644	0.707	0.643

(b) Fast Motion.

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.797	0.714	0.691	0.641	0.612	0.653	0.731	0.641
mPN	0.787	0.719	0.670	0.641	0.620	0.679	0.735	0.696

(c) In-Plane Rotation.

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.856	0.806	0.786	0.691	0.665	0.699	0.704	0.608
mPN	0.847	0.787	0.747	0.681	0.664	0.719	0.704	0.674

(d) Illumination Variation.

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.830	0.650	0.596	0.631	0.420	0.461	0.777	0.293
mPN	0.801	0.648	0.635	0.615	0.549	0.572	0.766	0.455

(e) Low Resolution.

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.850	0.742	0.778	0.624	0.562	0.622	0.703	0.638
mPN	0.833	0.736	0.754	0.617	0.548	0.639	0.699	0.641

(f) Motion Blur.

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.804	0.722	0.742	0.646	0.573	0.636	0.696	0.612
mPN	0.793	0.717	0.715	0.643	0.567	0.654	0.687	0.661

(g) Out-of-Plane Rotation.

	fdKCF*	BACF	ECO-HC	MKCFup	DSST	Staple	CFNet	LCT
mOP	0.804	0.704	0.725	0.588	0.551	0.607	0.687	0.475
mPN	0.824	0.716	0.730	0.608	0.594	0.676	0.703	0.596

(h) Scale Variation.

Table 2: The mean OP (mOP) and mean P_{norm}@0.2 (mPN) of our fdKCF* and seven modern real-time CF trackers on eight representative attributes of OTB-2015. The best three results are shown in red, blue, and magenta, respectively. fdKCF* outperforms all other real-time CF trackers by large margins on all attributes.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.752	0.773	0.732	0.742	0.670	0.667	0.660	0.604
mPN	0.767	0.753	0.738	0.741	0.670	0.651	0.664	0.680

(a) Deformation.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.834	0.808	0.726	0.796	0.735	0.718	0.712	0.669
mPN	0.798	0.778	0.724	0.777	0.709	0.694	0.701	0.740

(b) Fast Motion.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.797	0.794	0.739	0.742	0.702	0.662	0.682	0.673
mPN	0.787	0.773	0.736	0.737	0.694	0.649	0.678	0.735

(c) In-Plane Rotation.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.856	0.856	0.784	0.805	0.719	0.742	0.771	0.632
mPN	0.847	0.836	0.783	0.779	0.702	0.708	0.753	0.727

(d) Illumination Variation.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.830	0.720	0.679	0.751	0.719	0.684	0.682	0.331
mPN	0.801	0.721	0.686	0.759	0.696	0.675	0.669	0.501

(e) Low Resolution.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.850	0.848	0.784	0.843	0.754	0.730	0.770	0.700
mPN	0.833	0.824	0.769	0.818	0.732	0.706	0.751	0.748

(f) Motion Blur.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.804	0.819	0.765	0.783	0.726	0.664	0.708	0.637
mPN	0.793	0.787	0.763	0.760	0.706	0.640	0.695	0.701

(g) Out-of-Plane Rotation.

	fdKCF*	ECO	GPRT	C-COT	deep	SRDCF	decon	HCF
mOP	0.804	0.808	0.734	0.788	0.732	0.670	0.737	0.518
mPN	0.824	0.807	0.760	0.797	0.743	0.671	0.745	0.670

(h) Scale Variation.

Table 3: The mean OP (mOP) and mean P_{norm}@0.2 (mPN) of our fdKCF* and seven modern CF trackers that produce state-of-the-art localization accuracy on eight representative attributes of OTB-2015. The best three results are shown in red, blue, and magenta, respectively. fdKCF* outperforms other state-of-the-art trackers on most attributes.

temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Transactions on Image Processing*, 2019.

[15] Zheng Zhu, Qiang Wang, Bo Li, Wei Wu, Junjie Yan, and

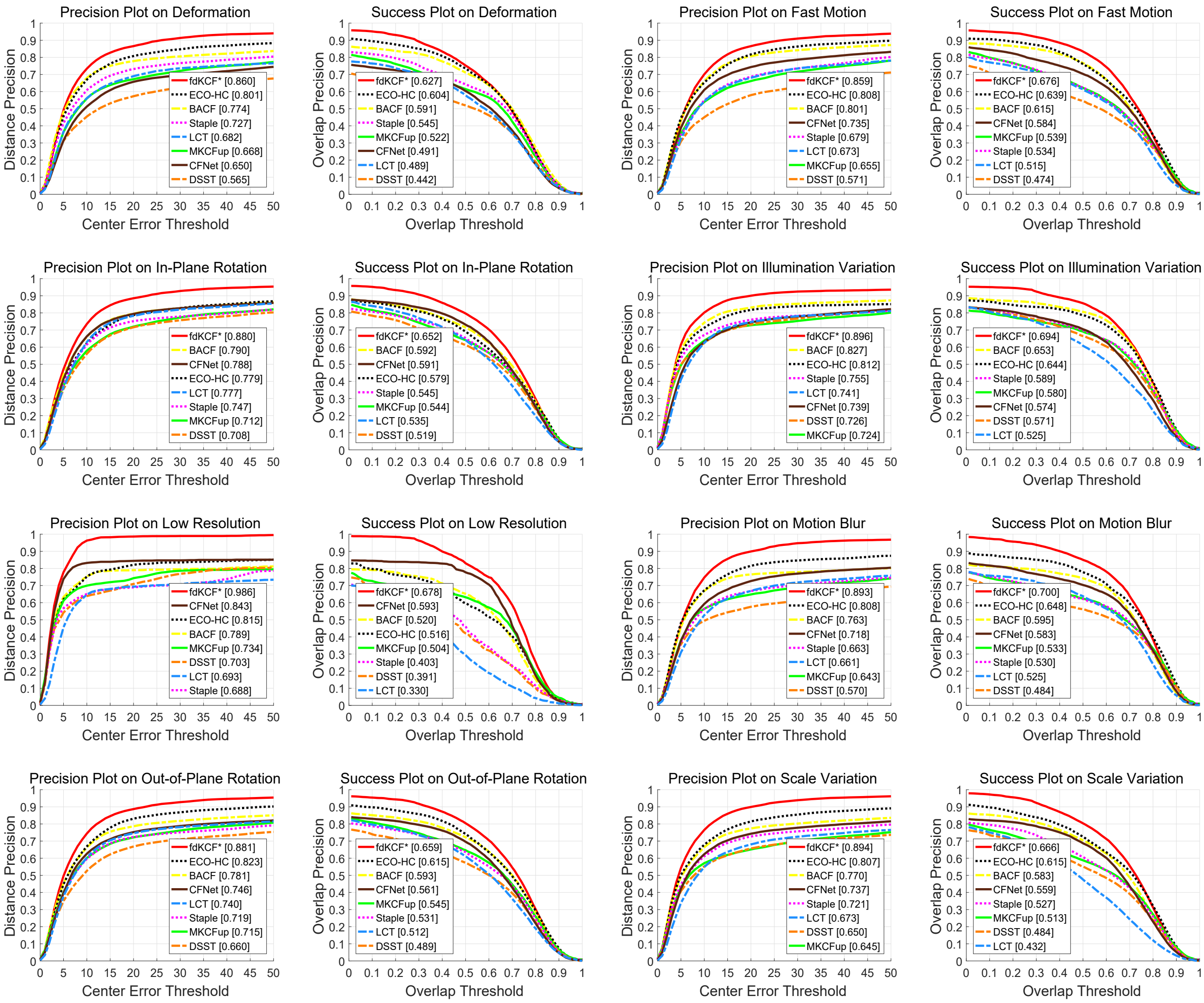

Figure 1: The mean precision and success plots of our fdKCF* and seven modern real-time CF trackers on eight representative attributes of OTB-2015. The mean distance precisions and AUCs are reported in the legends. fdKCF* outperforms all other real-time CF trackers by large margins on all attributes.

Weiming Hu. Distractor-aware siamese networks for visual object tracking. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 101–117, 2018.

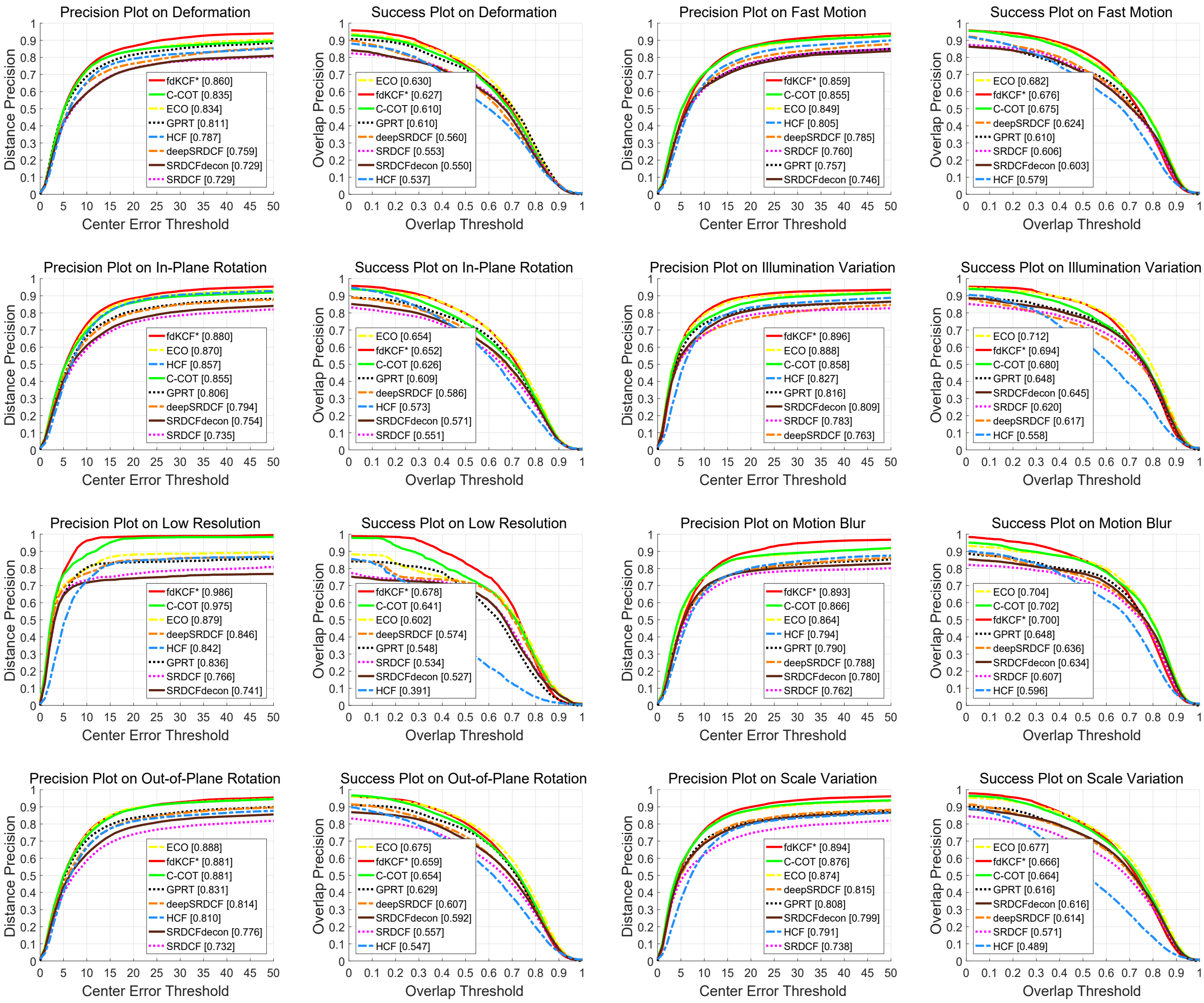

Figure 2: The mean precision and success plots of our fdKCF* and seven modern CF trackers that produce state-of-the-art localization accuracy on eight representative attributes of OTB-2015. The mean distance precisions and AUCs are reported in the legends. On the whole, fdKCF* is competitive with ECO and remarkably superior to other state-of-the-art CF trackers.

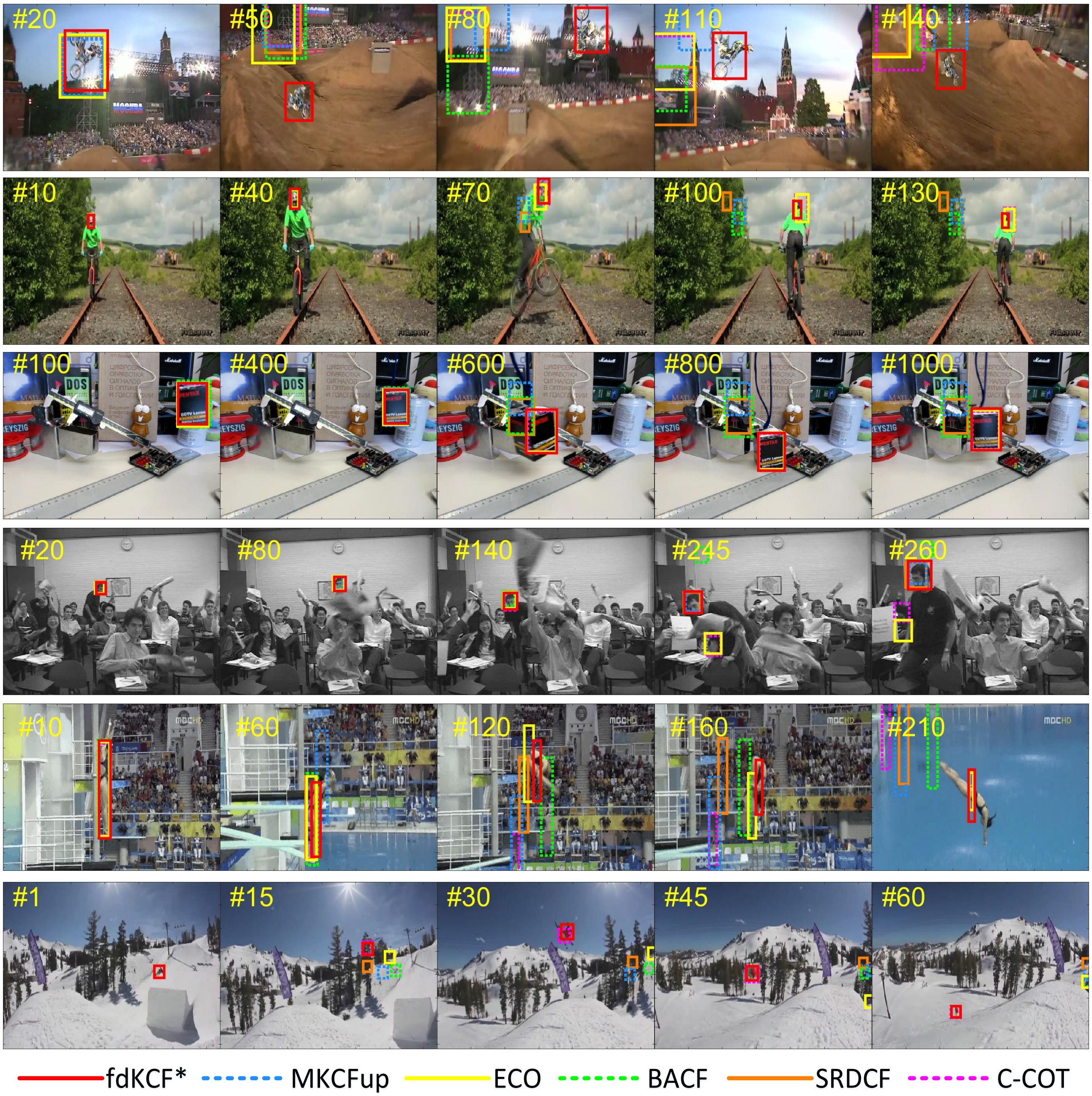

Figure 3: Qualitative comparison of our fdKCF* and five representative CF trackers, MKCFup, ECO, BACF, SRDCF and C-COT, on six hard and very hard sequences [6] of OTB-2015. They are MotorRolling, Biker, Box, Freeman4, Diving, and Skiing from the top down. Our fdKCF* is superior to other trackers.