A Perfect Tracker on CVPR TB-50 & ALOV300++

Bolun Cai
School of Electronic and Information Engineering
South China University of Technology, Guangzhou, China
caibolun@gmail.com

***Abstract*—Visual object tracking is one of the fundamental problems of computer vision, with wide-ranging applications including video surveillance, human-machine interfaces and robot perception. Although visual tracking has been investigated intensively in the past decade, it is still an enormous challenge in real application because of various factors such as pose, occlusion, scale and illumination. Recent tracking algorithms can be split into two main modules generally: feature extraction and tracking model. This report shows a perfect online tracker, which was extensively evaluated on the CVPR 2013 Tracking Benchmark (TB-50) including 50 sequences and the Amsterdam Library of Ordinary Videos (ALOV300++) dataset. The experimental results demonstrated the superior performance of it in comparison with other state-of-art trackers.**

***Index Terms*—perfect tracker, TB-50, ALOV300++;**

I. EXPERIMENTAL RESULTS

The proposed tracker is implemented in MATLAB 2013A on a PC with Intel Core2 CPU (2.66 GHz) with 2 GB memory, and runs about 50 frames per second (fps) in this platform.

A. CVPR TB-50

We compared the proposed method with 10 state-of-the-art trackers (TGPR[1], KCF[2], Struck[3], SCM[4], TLD[5], CXT[6], VTD[7], VTS[8], CSK[9], ASLA[10], LOT[11], OAB[12]) on the CVPR2013 benchmark [13] that includes 50 sequences showed in Fig. 1. Each sequence is tagged with a number of attributes indicating to the presence of 11 different challenges, including Illumination Variation (IV), Scale Variation (SC), Occlusion (OCC), Deformation (DEF), Motion Blur (MB), Fast Motion (FM), In-Plane Rotation (IPR), Out-of-Plane Rotation (OPR), Out-of-View (OV), Background Clutters (BC), Low Resolution (LR). The best way to evaluate trackers is still a debatable subject. Averaged measures like mean center location error or average bounding box overlap penalize an accurate tracker that fails for short-time more than an inaccurate tracker. According to [13], the precision plot shows the percentage of frames on which the Center Location Error (CLE) of a tracker is within a given threshold e, where CLE is defined as the center distance between tracker output $(\hat{x}, \hat{y})$ and ground truth (x_g, y_g).

Fig.2 shows the precision plots containing the mean error over all the 50 sequences, and a representative precision score ($e = 20$) is used for ranking. In the precision plot, the proposed tracker outperforms TGPR[3] by 4.9% in mean CLE at the threshold of 20 pixels. On the other hand, the tracking drift of our tracker is less than the baseline KCF in the high-precision ($e < 20$), and our tracker also acquires the more high accuracy than KCF.

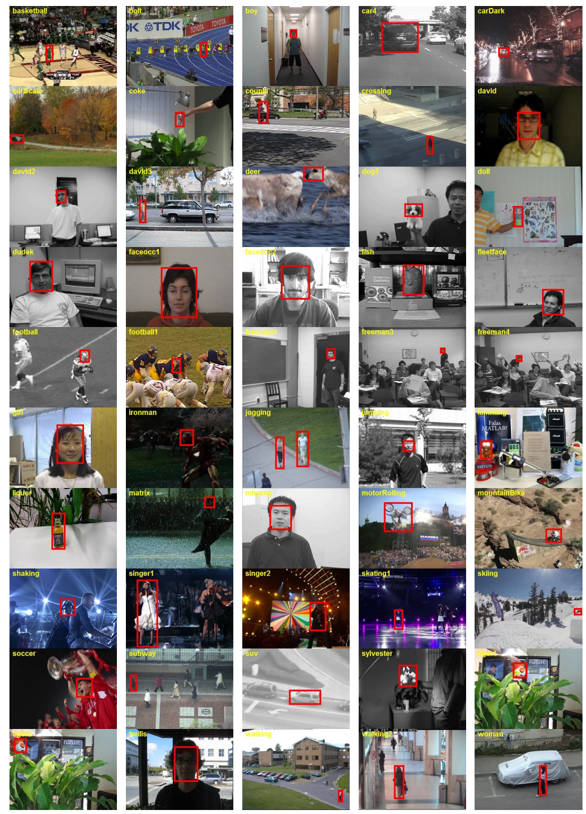

Fig. 1. CVPR 2013 Tracking Benchmark (TB-50) for evaluation

B. ALOV++

To further validate the robustness of our tracker, we conducted the second evaluation on a larger dataset [14], namely Amsterdam Library of Ordinary Videos (ALOV300++) showed as Fig. 3, which is recently developed by Smeulders et al.. It consists of 14 challenge subsets, totally 315 sequences and focuses on systematically and experimentally evaluating trackers robustnesses in a large variety of situations including illuminations, transparency, specularity, confusion with similar objects, clutter, occlusion, zoom, severe shape changes, motion patterns, low contrast, and so on. In [14], survival curves based

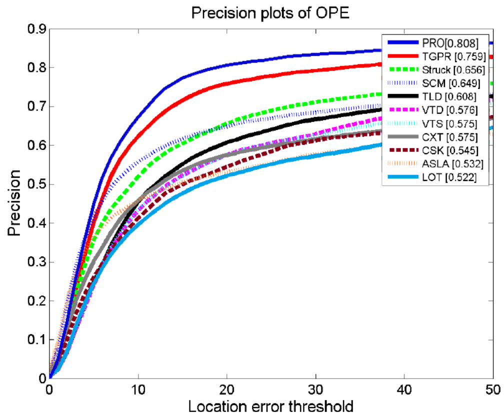

Fig. 2. Precisions plots comparing with state-of-the-art trackers

Fig. 3. Amsterdam Library of Ordinary Videos (ALOV300++) for evaluation

on F-score were proposed to evaluate trackers robustnesses and demonstrated its effectiveness. To obtain the survival curve of a tracker, a F-score for each video is computed as $F = 2 \times (precision \times recall)/(precision + recall)$, where $precision = ntp/(ntp + nfp)$, $recall = ntp/(ntp + nfn)$, and ntp, nfp, nfn respectively denote the number of true positives, false positives and false negatives in a video. A survival curve shows the performance of a tracker on all videos in the dataset. The videos are sorted according to the F-score. By sorting the videos, the graph gives a birds eye view in cumulative rendition of the quality of the tracker on the whole dataset.

To evaluate our tracker on ALOV300++ dataset, We compare our tracker with 19 popular trackers that were evaluated in [14]. In addition, we also ran MEEM[15] on ALOV++, which ranks the second best in the previous evaluation. The survival curves of the top ten trackers and the average F-scores over all sequences are shown in Figure 4, which demonstrates that our tracker achieves the best overall performance over 21 compared trackers in this comparison. The average F-score of our tracker on ALOV300++ is 0.74, which is significantly better than Struck (0.66)[3], MEEM (0.65)[15], TLD (0.61)[5] and the other competitors.

II. Conclusion

In this work, we demonstrate that it is possible to build a perfect model to track targets successfully, which achieves excellent result in complicated and diverse environments. The experimental results on two large datasets demonstrate that the proposed tracker is capable of taking advantage of both the short-term and long-term systems and significantly boosting the tracking performance.

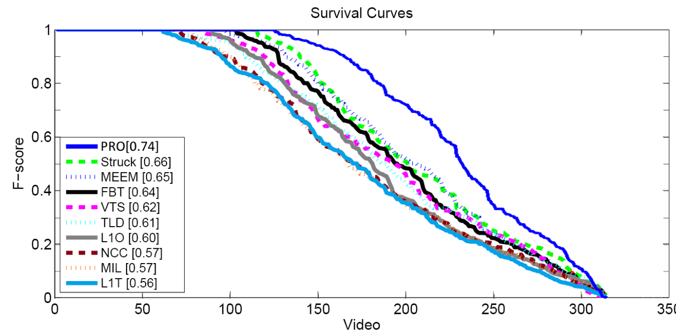

Fig. 4. The survival curves for top ten trackers on AlOV++ dataset

III. Acknowledgment

This work is supported by the National Natural Science Foundation of China (#61171142, #61401163), the Science and technology Planning Project of Guangdong Province of China (#2011A010801005, #2012B061700102).

References

[1] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 188–203.

[2] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," 2014.

[3] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," in *International Conference on Computer Vision*. IEEE, 2011, pp. 263–270.

[4] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1838–1845.

[5] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-n learning: Bootstrapping binary classifiers by structural constraints," in *Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 49–56.

[6] T. B. Dinh, N. Vo, and G. Medioni, "Context tracker: Exploring supporters and distracters in unconstrained environments," in *Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 1177–1184.

[7] J. Kwon and K. M. Lee, "Visual tracking decomposition," in *Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 1269–1276.

[8] J. Kwon and L. K. Mu, "Tracking by sampling trackers," in *ICCV*. IEEE, 2011, pp. 1195–1202.

[9] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *ECCV*. Springer, 2012, pp. 702–715.

[10] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1822–1829.

[11] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," in *Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1940–1947.

[12] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *ECCV*. Springer, 2008, pp. 234–247.

[13] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Computer Vision and Pattern Recognition*. IEEE, 2013, pp. 2411–2418.

[14] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 7, pp. 1442–1468, 2014.

[15] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 188–203.