International Journal of Advanced Smart Convergence Vol.4 No.2 20-28 (2015)
http://dx.doi.org/10.7236/IJASC.2015.4.2.20

IJASC 15-2-4

POSE-VIWEPOINT ADAPTIVE OBJECT TRACKING VIA ONLINE LEARNING APPROACH

Vinayagam Mariappan[*], Hyung-O Kim[**], Minwoo Lee[**], Juphil Cho[***], Jaesang Cha[**†]

[*]*Media IT Engineering, Seoul National Univ., of Science and Tech., Seoul, Korea*
[**]*Graduate School of NID Fusion Tech., Seoul National Univ., of Science and Tech., Seoul, Korea*
[***]*Dept. Of Integrated IT & Communication Eng., Kunsan National Univ., Kunsan, Korea*

Abstract

In this paper, we propose an effective tracking algorithm with an appearance model based on features extracted from a video frame with posture variation and camera view point adaptation by employing the non-adaptive random projections that preserve the structure of the image feature space of objects. The existing online tracking algorithms update models with features from recent video frames and the numerous issues remain to be addressed despite on the improvement in tracking. The data-dependent adaptive appearance models often encounter the drift problems because the online algorithms does not get the required amount of data for online learning. So, we propose an effective tracking algorithm with an appearance model based on features extracted from a video frame.

Keywords: *Object Tracking, Features Extraction, Effective Tracking Algorithm, Computer Vision, Target Detection and Tracking, Online Learning, TLD, Random Forest, Naive Bayes, long-Term Object Tracking, Adaptive Appearance Model.*

1. INTRODUCTION

Unmanned Aerial Vehicles (UAV) have been an active filed for research and play a major role in several scenarios, such as surveillance, environment conservation, industrial inspection, media shootings and disaster management. Furthermore, with the availability of low cost, robust and small video cameras, UAV video has been one of the fastest growing data sources in the last couple of years. Some of the most recent work within UAV field include autonomous see-and-avoid systems and autonomous visual based landing.

Nowadays, secure autonomous motion control during the whole flight is essential for wide spread acceptance of UAV. Moreover, autonomous take-off and landing is a necessary capability for UAV operating. In addition, the landing procedure is the most critical phase during the entire UAV flight. However, navigating based on Global Navigation Satellite System (GNSS) is not sufficient because of multipath reception and jamming. To deal with these problems, vision-based UAV control systems have been proposed.

Manuscript Received: Sept. 5, 2015 / Revised: Sept. 30, 2015 / Accepted: Oct. 15, 2015
Corresponding Author: chajs@seoultech.ac.kr
Tel: +82-2-970-6431
Graduate School of NID Fusion Tech., Seoul National Univ., of Science and Tech., Seoul, Korea

Due to the size, weight and power constrained of UAV, the initial solution was a common approach to object detection and tracking in UAVs is to send the recorded video data to a ground station for processing. Object detection and tracking is then performed at a high-end desktop computer, before command signals are sent to the autopilot control module located on-board the UAV. There are several complications associated with this approach like a reliable and fast wireless data connection is required between the UAV and the ground station at all times. If the UAV moves too far away from the ground station, the video signal is usually either transmitted with a huge lag time or worse, the data received at the ground station may be corrupted. This effectively limits the operational range of this approach drastically, and it is therefore in many cases not ideal.

Now, in recent years, computer hardware has become smaller, lighter, more power efficient and more powerful. This has lead to the possibility of implementing real-time object tracking directly on-board the UAV. In this paper, we propose an onboard autonomous tracking included UAV system. The main challenges of visual tracking can be attributed to the difficulty in handling appearance variability of a target object. Intrinsic appearance variabilities include pose variation and shape deformation of a target object, whereas extrinsic illumination change, camera motion, camera viewpoint, and occlusions inevitably cause large appearance variation. Due to the nature of the tracking problem, it is imperative for a tracking algorithm to model such appearance variation. Our method addresses these issues with an adaptive approach combining an online-learning learning based on adaptive appearance model feature update on detector to enhance tracking efficiency.

2. RELATED WORKS

Object tracking is relatively easy for humans. Humans respond quickly to visual information by recognizing temporal consistency and memorizing useful visual features to recover from tracking failures when the target leaves field of- view. Memory is one of the most powerful, but least well understood, functions of the human brain. With the sophisticated memory system, humans are capable of adapting to complex environments and behaving stably and consistently when facing temporal issues.

Most current target detection and tracking methods can be divided into the following several categories: (1) Statistical model-based algorithms[25-26]: firstly, a large-scale data are trained to achieve the distribution information of targets; then, the distance between targets are calculated to obtain the number of matched features. Such algorithms mainly apply for the scene with a single background. (2) Knowledge-based algorithms [27-28]: such algorithms can solve the limitations of statistical model-based algorithms, but it also introduces some more problems needed to be redefined in a new scene, including verification difficulties, large costs. (3) Model-based algorithms[29-30]: such algorithms first extract target features; then, construct spatial model of targets; next time, utilize selected features to initialize system parameters; finally, predict the location of targets via the characteristics of targets. (4) Neural networks-based and expert systems-based algorithms [31-32]: such algorithms can solve the problems which conventional algorithm cannot solve. However, the real-time of such algorithms is so poor.

Various approaches that form the basis of existing trackers can be used to model the memory of the target appearance. In [6], Ross et al. proposed to incrementally learn a low-dimensional subspace of the target representation. Later, Mei et al. [7] introduced sparse representations for tracking, subsequently adopted in many trackers [8, 9], in which the memory of the target appearance is modeled using a small set of target

instances. In contrast to the generative approaches used in [10] and [11], discriminative methods [12, 13, 14, 15, 16, 17] have been proposed that consider both foreground and background information. In particular, Struck [15] is one of the best performing trackers and has been highlighted in several recent studies [18, 43, 19]. In [15], Hare et al. introduced structured SVM for tracking and trained a classifier using samples with structured labels. The correlation filter-based trackers [20, 21, 22, 16, 23, 24] are becoming increasingly popular due to their promising performance and computational efficiency. However, most of these trackers depend on the spatiotemporal consistency of visual cues and adopt relatively risky update schemes; therefore, they can only handle short-term tracking.

Online object tracking has long been a popular topic in computer vision in current days. A large number of trackers have been proposed [5, 4], and the recent publication of benchmark datasets containing large numbers of sequences and standardized quantitative evaluation metrics is accelerating the pace of development in this field [3, 2, 1]. After analyzing theoretical principles and implementation mechanisms of the TLD algorithm, this paper presents a method to improve the performance of the TLD algorithm. TLD algorithm is a single-target algorithm which can track target for a long time. The advantage of the TLD algorithm can be summarized as the following two aspects: (1) combinating detection with tracking to deal with the issue of missing or deformed; (2) utilizing an improved semi-supervised learning method to update the detection and tracking module to enhance the performance of stability, robustness, and reliability.

This paper first analyzes the principle of adapting appearance based model and makes improvement of it; then, an improved Adaptive Appearance Model algorithm is utilized to enhance the reliability of TLD algorithm; finally, the predicted results obtained by the Adaptive Appearance Model are utilized to narrow the detection area obtained by the TLD algorithm to further improve the real-time of target tracking.

3. ONLINE LEARNING

TLD (Tracking-Learning-Detection) [1] algorithm is a famous online learning tracking algorithm. It is a kind of discriminative approach and is often used to process long-term video stream.

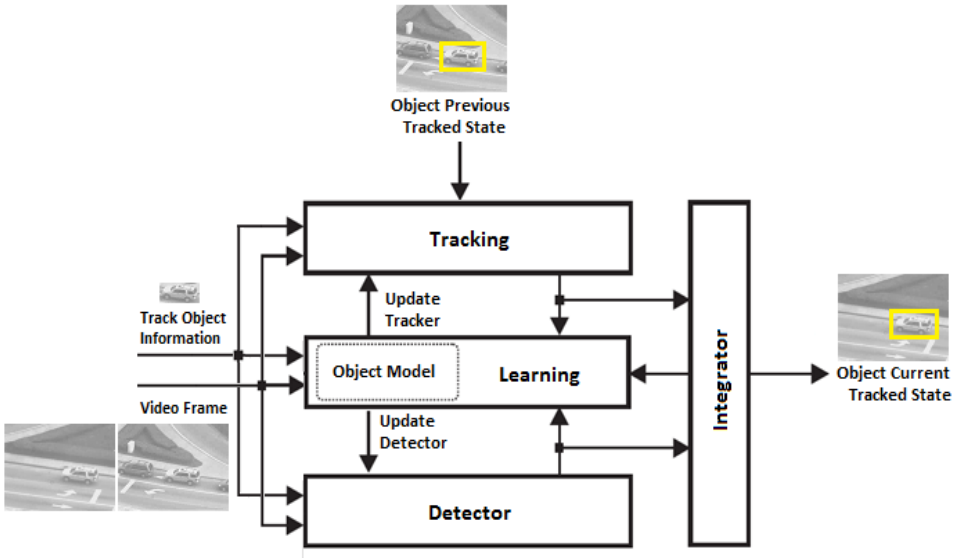

Figure 1. The Block Diagram of TLD Framework

Figure 1 shows the implement process of TLD. We can see TLD contains three main function parts: tracking, learning, and detection. Tracking and detection influence and correct each other. Subsequently, TLD uses labeled and unlabeled samples for the detector classifier learning. A few candidates from the detection part are integrated with the bounding box from the tracking part into a single target region in current frame, making the approach more robust than other trackers.

I. TRACKING

The tracking part adopts Median-Flow tracker to estimate the target region between consecutive frames under the assumption that the frame-to-frame motion is limited and the appearance of object changes little, producing a candidate target window in the current frame. Median-Flow tracker [8] uses twice times pyramidal Lucas-Kanade tracker (LKT) [9] to estimate feature points within the last bounding box which sample uniformly in the window, then selects the points whose distance error is smaller than the median distance error and local similarity is bigger than the median similarity. When the number of fitted feature points meets the requirements and median distance error satisfies the thresholds respectively, the tracking result is valid and the tracker outputs the only bounding box. Tracking makes an estimation based on target position in the last frame. However, it is easy to drift when comes out occlusion.

II. DETECTION

Detection part processes every frame independently. The detector adopts the cascaded classifiers structure. The patches generated by sliding window get through three classifiers successively. The three-stage classifiers are patch variance, ensemble classifier, and nearest neighbor in proper order. Patch variance rejects patches by the rule that gray value variance must be bigger than threshold which is set by initial bounding box in first frame. Ensemble classifier uses the comparison of gray value between the vertical and horizontal distribute pairs of pixels as 2bitBP feature getting through the fast Random Ferns. For nearest neighbor classifier, relative similarity (Sr) of candidate window which stands for the similarity confident with appearance model must exceed the threshold. Detection part bases on the appearance model and searches the match candidate patches in whole image. Hence, detector can correct the tracker or re-initialing the tracker.

III. LEARNING

TLD proposed the P-N Learning for collecting the samples and update the appearance model. Based on the initial target bounding box in first frame, the detector learns a primitive classifier and a primitive appearance model coming from the shifted good boxes around the best patch and the bad boxes far away from the best patch. In the followed frame, once the target is localized, positive samples are selected in and around the target and negative samples selected at a distance, and then P-N experts classify the false positive and false negative to retrain detector classifier and update target appearance model when the conservative similarity of target meets some conditions. In current frame, integrator outputs the candidate patch as new target's position. The result of tracker holds big weight on the ultimate result, however, if there are some candidates from detector which have better relative similarity than that of tracker, these candidates will dominant the ultimate result and the tracker will reinitialize. If neither detector nor tracker outputs a patch, TLD declares loss of target.

For the perfect cooperation with every part, TLD can be used to track an unknown object. It has perfect

performance when faces with partial occlusion, scale variation and motion blur. With the help of good learning mechanism, TLD can update the appearance model and classifiers when the target appearance changes slowly and coherently. However, we can find 2bitBP feature is the gray-value comparison and sliding window used as search mechanism is low-efficiency, in real scene the target object usual has a stable color distribute even though the object's gesture changes. The adaptive appearance object model based on the color feature can make the tracking result more precise. Based on the loss of the use of color feature appearance on the object, this paper propose the TLD with the adaptive appearance model feature.

4. ADAPTIVE APPEARANCE MODEL

The Real-Time UAV capture frames by nature it may possess more of the occlusion and disappearance of the object, appearance and viewpoint changes, object scale changes, illumination changes, noise in image noise due to capturing in moving and vibration on the system. This nature affects the performance of the object tracking more precisely. To overcome the performance of tracking, the TLD algorithm need to be tuned on object recognition by extracting appropriate object futures from video frames. So this paper proposes modified TLD using Adaptive Appearance Model.

Color feature is regarded as an important evident to visual image processing. In visual tracking, color tracking has been applied in some real examples. Good performance supports us to solve the weakness of TLD by using color feature bases appearance model. There are so many kinds of color representations at present, like RGB, YCbCr, HSV, HUE, Opp-Angle. And different color representation has different sensibility on the illumination and hue and then affects the performance of image processing, so the choice of color feature representation is important for visual tracking algorithm.

The Color based appearance model gives the perfect performance on the object recognition, segmentation and visual tracking applications, so this paper proposes the color based appearance model as the color feature representation to classify the target and improve the TLD. The proposed method block diagram is shown in Figure 2. The Appearance Model estimator initialized with the object need to be tracked and it estimated the color based appearance parameters and initialize the TLD Tracking module with the appearance features. The Tracking module takes the fusion future from color based appearance model and grey scaled based object model to make decision on object need to be tracked.

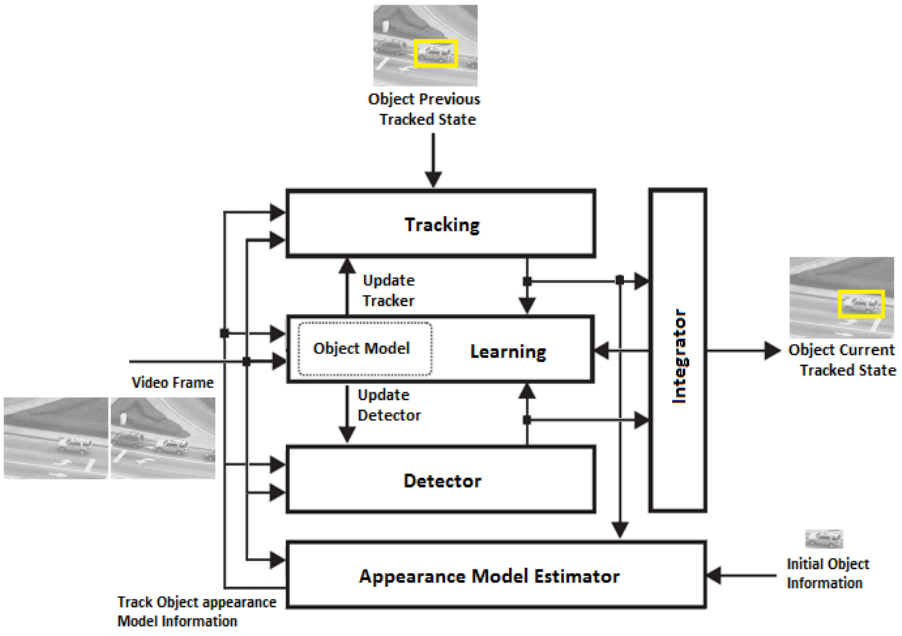

Figure 2. The TLD with Adaptive Appearance Model Block Diagram

6. RESULT AND ANALYSIS

The proposed approach is implemented on Visual Studio 2013 on Intel i7 Core with Windows 7 platform using OpenCV3.0 and OpenTLD. The experiment results demonstrate that the proposed adaptive appearance based TLD algorithm can detect and track targets accurately under complex circumstances, such as disappear, turning, and mutual occlusion. The implementation is analyzed and verified with tracking video test files like pedestrian.mpg, volkswagen.mpg, car.mpg. Also the algorithm is verified on live video stream from camera on WiFi network using HTTP and RTP/RTSP live streams.

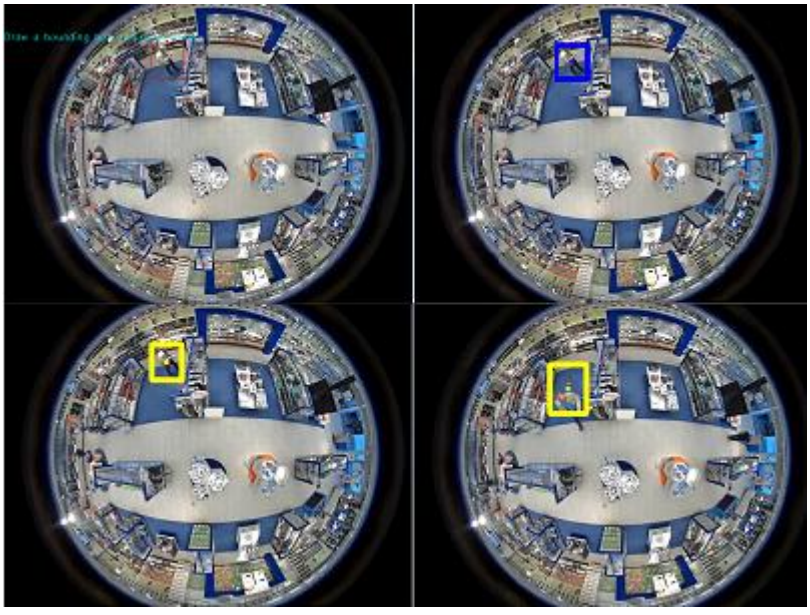

Figure 3. The Proposed Approach Implementation Result

7. CONCLUSIONS

In this paper, to improve the tracking performance of the TLD algorithm, we use color based object appearance model representation. The original TLD uses only the gray information to classify the candidate patches to detect the objects; however, the gray information is easy to be effected by illumination variance and has low discrimination power. Compared with the gray image, Color based representation has high discriminative power and partial photometrical invariance so that it can locate target more precisely. From our experiment, we can see color based object appearance model can promote the precision rate and success rate under different evaluation model in different scenes in real-time.

ACKNOWLEDGEMENT

This research was supported by the ICT R&D program of MSIPITlTP. [A broadcasting vehicle-mountable multi copter CAM system]

REFERENCES

[1] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," In CVPR, pp. 2411–2418, 2013.
[2] A. Smeulders, D. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *TPAMI*, Vol. 36, No. 7, pp.1442–1468, 2014.
[3] M. Kristan and R. Pflugfelder et al. "The visual object tracking VOT2014 challenge results," In ECCV Workshop, pp.1–27, 2014.
[4] H. Yang, L. Shao, F. Zheng, L. Wang, and Z. Song, "Recent advances and trends in visual tracking: A review," *Neurocomputing*, Vol. 74, No. 18, pp. 3823–3831, 2011.
[5] X. Li,W. Hu, C. Shen, Z. Zhang, A. Dick, and A. V. D. Hengel, "A survey of appearance models in visual object tracking," *TIST*, Vol. 4, No. 4, pp. 58, 2013.
[6] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *IJCV*, Vol. 77(1-3), pp. 125–141, 2008.
[7] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *TPAMI*, Vol. 33, No. 11, pp. 2259– 2272, 2011.
[8] Z. Hong, X. Mei, D. Prokhorov, and D. Tao, "Tracking via robust multi-task multi-view joint sparse representation," In ICCV, pp. 649–656, 2013.
[9] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," In CVPR, pp. 2042–2049, 2012.
[10] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *IJCV*, Vol. 77(1-3), pp. 125–141, 2008.
[11] G. Nebehay and R. Pflugfelder, "Consensus-based matching and tracking of keypoints for object tracking," In WACV, pp. 862–869, 2014.
[12] S. Avidan, "Support vector tracking," *TPAMI*, Vol. 26(8), pp. 1064–1072, 2004.
[13] S. Avidan, "Ensemble tracking," *TPAMI*, Vol. 29(2), pp. 261–271, 2007.
[14] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *TPAMI*, Vol. 33(8), pp. 1619–1632, 2011.
[15] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," In ICCV, pp.

263–270, 2011.

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

[17] Z. Hong, X. Mei, and D. Tao, "Dual-force metric learning for robust distracter-resistant tracker," In ECCV, pp. 513–527, 2012.

[18] Y. Pang and H. Ling, "Finding the best from the second bestsinhibiting subjective bias in evaluation of visual tracking algorithms," In ICCV, pp. 2784–2791, 2013.

[19] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," In CVPR, pp. 2411–2418, 2013.

[20] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," In CVPR, pp. 2544–2550, 2010.

[21] M. Danelljan, G. H¨ager, F. S. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," In BMVC, 2014.

[22] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer, "Adaptive color attributes for real-time visual tracking," In CVPR, pp. 1090–1097, 2014.

[23] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "Highspeed tracking with kernelized correlation filters," *TPAMI*, pp. 583–596, 2015.

[24] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," In ECCV Workshop, 2014.

[25] J. Wu, G. Li, and F. Ma, "Research on target tracking algorithm using improved current statistical model," International Conference on Electrical and Control Engineering, pp. 2515-2517, 2011.

[26] C. Huang, P. Feng, L. Cao, H. Huang, and H. Cheng, "A target tracking algorithm based on current statistical model for adjusting acceleration variance of maneuver target," *Journal of Northwestern Polytechnical University*, 2014.

[27] Q. Li, L. Kong, and X. Yang, "The knowledge-based tracking using geographic information," IEEE Conference on Radar, pp. 777-780, 2011.

[28] A. Mazinan, A. Amir, and M. Kazemi, "A knowledge-based objects tracking algorithm in color video using Kalman filter approach," IEEE Conference on Information Retrieval and Knowledge Management, 2012.

[29] I. Barkana, "On adaptive model tracking with mitigated passivity conditions," Israel Annual Conference on Aerospace Sciences, pp. 512-541, 2012.

[30] J. Hou, X Li, and Z. Jing, "Multiple model tracking of manoeuvring targets accounting for standoff jamming information," *IET Radar, Sonar and Navigation*, Vol. 7, No. 4, pp. 342-350, 2013.

[31] W. Kazimierski and A. Stateczny, "Optimization of multiple model neural tracking filter for marine targets," International Radar Symposium, pp. 543-548, 2012.

[32] H. Wang, B. Chen, X. Liu, K. Liu, and C. Lin, "Adaptive neural tracking control for stochastic nonlinear strict-feedback systems with unknown input saturation," *Information Sciences*, Vol. 269, pp. 300-315, 2014.

[33] S. Avidan, "Ensemble tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(2), pp. 261–271, Feb. 2007.

[34] B. Babenko, Ming-Hsuan Yang, and S. Belongie, "Visual tracking with online multiple instance learning," In 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), pp. 983–990, IEEE, June 2009.

[35] R. Brunelli, *Template Matching Techniques in Computer Vision: Theory and Practice, Wiley Publishing*,

2009.

[36] Z. Kalal, J. Matas, and K. Mikolajczyk, "Online learning of robust object detectors during unstable tracking," In Proceedings of the IEEE On-line Learning for Computer Vision Workshop, pp. 1417–1424, 2009.

[37] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints," In 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 49–56, IEEE, June 2010.

[38] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-Backward Error: Automatic Detection of Tracking Failures," In International Conference on Pattern Recognition, pp. 23–26, 2010.