In Defense of Color-based Model-free Tracking

Horst Possegger, Thomas Mauthner, Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology

We address the problem of model-free online object tracking based on color representations. According to recent benchmark evaluations, such trackers often tend to drift towards regions which exhibit a similar appearance compared to the object of interest. To overcome this limitation, we propose an efficient discriminative object model which allows us to identify potentially distracting regions in advance. We exploit this knowledge to adapt the object representation beforehand so that distractors are suppressed and the risk of drifting is significantly reduced. In addition to its favorable simplicity, this representation also enables accurate scale estimation, as shown in Figure 1.

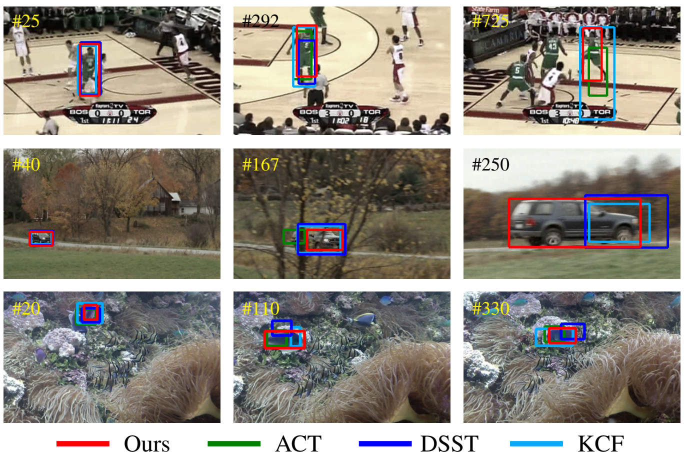

Figure 1: Sample results of our approach and recent state-of-the-art trackers.

Over the last decade the research focus has shifted to trackers based on well engineered features such as HOG [3], correlation filters [1], and more complex color features [2]. In contrast to this development, we show that trackers based on standard color representations (*i.e.* color histograms) can still achieve state-of-the-art performance.

We base our tracking approach on two primary requirements for online model-free trackers: First, considering subsequent frames, useful object models must be able to distinguish the object from its current surrounding background. Second, to reduce the risk of drifting towards regions which exhibit similar appearance at a future time step, such distracting regions must be identified beforehand and should be suppressed to ensure robustness.

To address the first requirement, *i.e.* distinguishing object pixels $\mathbf{x} \in \mathcal{O}$ from surrounding background pixels, we employ a color histogram based Bayes classifier on the input image I. Let $H_{\Omega}^{I}(b)$ denote the b-th bin of the non-normalized histogram H computed over the region $\Omega \in I$. Additionally, let $b_{\mathbf{x}}$ denote the bin b assigned to the color components of $I(\mathbf{x})$. Given a rectangular object region O (*i.e.* initial bounding box annotation or current tracker hypothesis) and its surrounding region S, we define the object-background model by applying Bayes rule as

$$P(\mathbf{x} \in \mathcal{O} | O, S, b_{\mathbf{x}}) = \begin{cases} \frac{H_O^I(b_{\mathbf{x}})}{H_O^I(b_{\mathbf{x}}) + H_S^I(b_{\mathbf{x}})} & \text{if } I(\mathbf{x}) \in I(O \cup S) \\ 0.5 & \text{otherwise,} \end{cases} \quad (1)$$

where unseen pixel values are assigned the maximum entropy prior of 0.5.

While this discriminative model distinguishes object and background pixels, regions which are visually similar to the object also yield high likelihood scores. However, we can exploit this information to identify and suppress such distracting regions in advance and thus, address the second tracking requirement. Whenever distracting regions D appear within a frame, we compute an object-distractor model $P(\mathbf{x} \in \mathcal{O} | O, D, b_{\mathbf{x}})$ by replacing the negative background samples $\mathbf{x} \in S$ in Eq. (1) with distracting pixels $\mathbf{x} \in D$.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

Combining both models, we obtain the final distractor-aware object model as $P(\mathbf{x} \in \mathcal{O} | b_{\mathbf{x}}) = \lambda_p P(\mathbf{x} \in \mathcal{O} | O, D, b_{\mathbf{x}}) + (1 - \lambda_p) P(\mathbf{x} \in \mathcal{O} | O, S, b_{\mathbf{x}})$, with the pre-defined weighting parameter λ_p.

Similar to recent state-of-the-art trackers, we first localize the object in a new frame and subsequently perform scale estimation. Following the tracking-by-detection principle, we propose an iterative non-maximum suppression strategy to localize the object. Using similarity scores based on visual appearance $s_v(\cdot)$ and Euclidean distance $s_d(\cdot)$, we find the current object location as $O_t^\star = \arg\max \left(s_v(O_{t,i}) s_d(O_{t,i}) \right)$. Additionally, our strategy yields potentially distracting regions D which are used to update the object model and thus reduce the risk of drifting. After localization, we segment the object via adaptive thresholding on $P(\mathbf{x} \in \mathcal{O} | b_{\mathbf{x}})$. This allows us to accurately estimate the object scale.

We evaluate our distractor-aware tracking approach on the recent Visual Object Tracking (VOT) benchmark datasets [4], VOT13 and VOT14. These datasets cover many challenging real-world scenarios, allowing us to both perform a fair comparison to the state-of-the-art and draw valid conclusions. Our evaluations show that the discriminative color model without distractor-awareness (*noDAT*) already achieves state-of-the-art accuracy, see Figure 2. By incorporating the distractor-aware representation (*DAT*), we can significantly increase the tracking robustness. Moreover, our scale-adaptive version (*DAT+scale*) outperforms the VOT challenge winner DSST [1]. Overall, our approach performs favorably compared with state-of-the-art approaches and enables robust online object tracking in real-time.

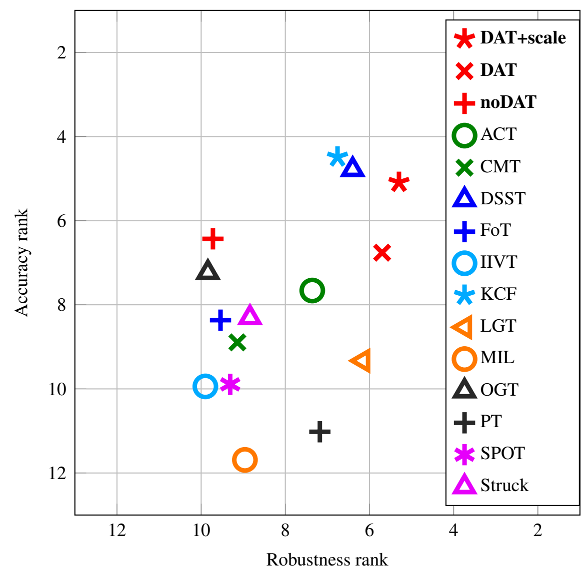

Figure 2: Comparison of our distractor-aware tracker (**DAT+scale** and **DAT**) with the baseline discriminative color model (**noDAT**) and several state-of-the-art approaches on the **VOT14** benchmark dataset. Top-performing trackers are located top-right.

[1] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg. Accurate Scale Estimation for Robust Visual Tracking. In *Proc. BMVC*, 2014.

[2] M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer. Adaptive Color Attributes for Real-Time Visual Tracking. In *Proc. CVPR*, 2014.

[3] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. High-Speed Tracking with Kernelized Correlation Filters. *PAMI*, 2015.

[4] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, L. Čehovin, G. Nebehay, T. Vojíř, G. Fernandez, et al. The Visual Object Tracking VOT2014 challenge results. In *Proc. VOT (ECCV Workshop)*, 2014.