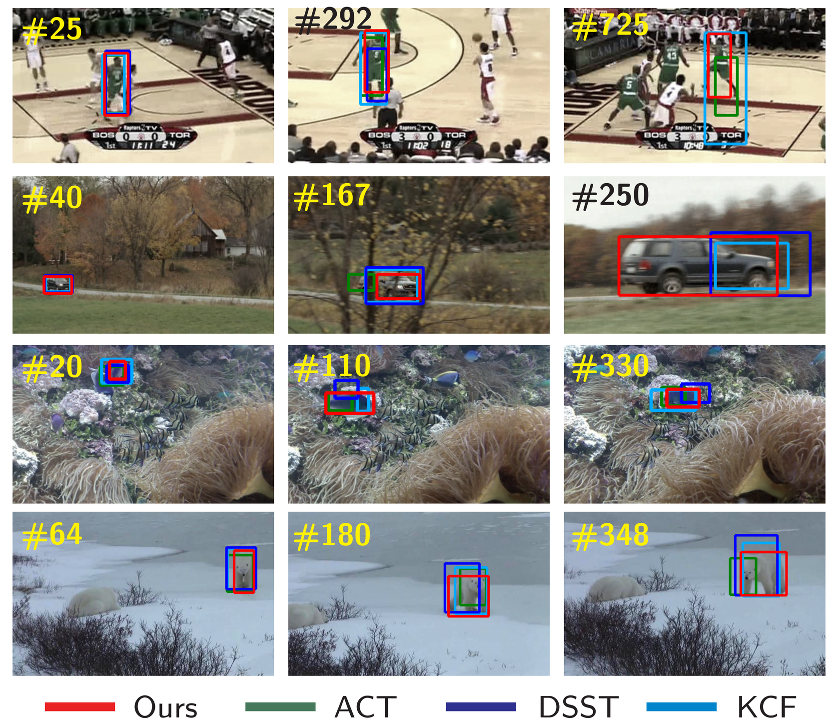
In Defense of Color-based Model-free Tracking

Horst Possegger Thomas Mauthner Horst Bischof

Graz University of Technology, Institute for Computer Graphics and Vision, Austria

Motivation and Contribution

- We address **color-based model-free online object tracking** where neither class-specific prior knowledge nor pre-learned object models are available.
- Recent benchmark evaluations (*e.g.* VOT [4]) show that **color-based trackers tend to drift** towards visually similar regions.
- **State-of-the-art** approaches rely on well engineered features (*e.g.* HOG [1]), correlation filters [3], and complex color features (*e.g.* color attributes [2]).
- We argue that trackers **based on standard color representations** still keep up with the state-of-the-art if they properly **address two key requirements**:
  - Distinguish the object from its surroundings.
  - Prevent drifting towards distracting regions.

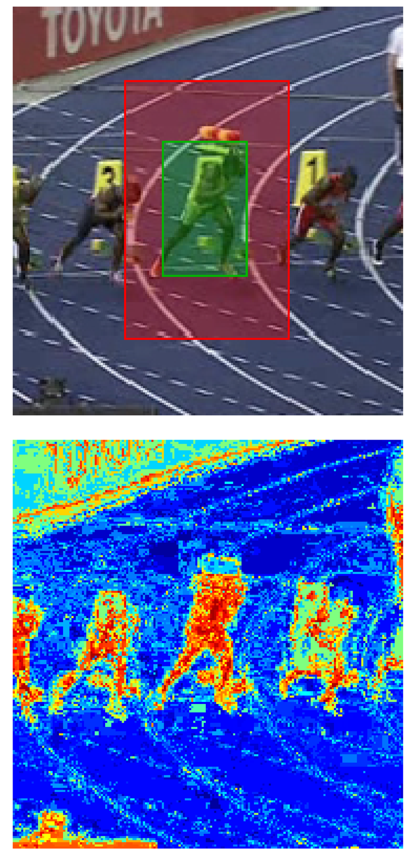


- To this end, we propose a **distractor-aware tracking approach** which addresses both requirements.
- **Supplemental material** publicly available (scan QR code).

Discriminative Object Model

- To **distinguish the object from its surrounding region**, we employ a Bayes classifier

$$P(\mathbf{x}\in\mathcal{O}\,|\,\mathbf{O},\mathbf{S},b_{\mathbf{x}}) \approx \frac{P(b_{\mathbf{x}}\,|\,\mathbf{x}\in\mathbf{O})\,P(\mathbf{x}\in\mathbf{O})}{\sum_{\Omega\in\{\mathbf{O},\mathbf{S}\}} P(b_{\mathbf{x}}\,|\,\mathbf{x}\in\Omega)\,P(\mathbf{x}\in\Omega)}.$$



- Color histograms $H^I_{\{\mathbf{O},\mathbf{S}\}}$ model the joint RGB distribution of image pixels $I(\mathbf{x})$ at location $\mathbf{x}$, where $b_{\mathbf{x}}$ denotes the corresponding bin

$$P(b_{\mathbf{x}}\,|\,\mathbf{x}\in\mathbf{O}) \approx \frac{H^I_{\mathbf{O}}(b_{\mathbf{x}})}{|\mathbf{O}|}, \qquad P(\mathbf{x}\in\mathbf{O}) \approx \frac{|\mathbf{O}|}{|\mathbf{O}|+|\mathbf{S}|},$$

$$P(b_{\mathbf{x}}\,|\,\mathbf{x}\in\mathbf{S}) \approx \frac{H^I_{\mathbf{S}}(b_{\mathbf{x}})}{|\mathbf{S}|}, \qquad P(\mathbf{x}\in\mathbf{S}) \approx \frac{|\mathbf{S}|}{|\mathbf{O}|+|\mathbf{S}|},$$

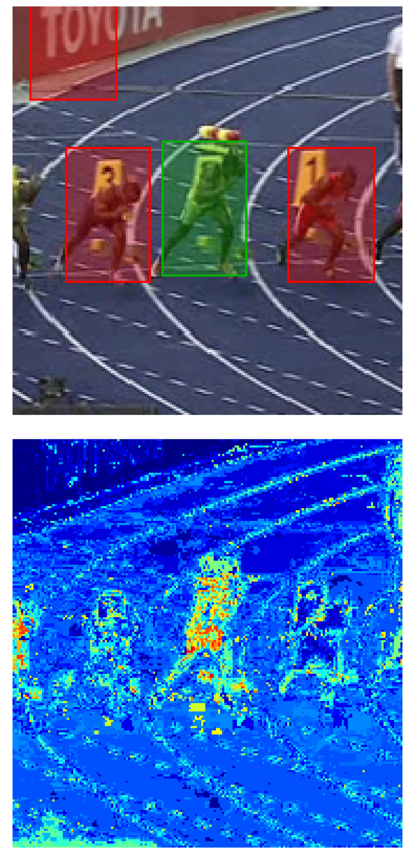
$$P(\mathbf{x}\in\mathcal{O}|\mathbf{O},\mathbf{S},b_{\mathbf{x}}) = \begin{cases} \frac{H^I_{\mathbf{O}}(b_{\mathbf{x}})}{H^I_{\mathbf{O}}(b_{\mathbf{x}})+H^I_{\mathbf{S}}(b_{\mathbf{x}})} & \text{if } I(\mathbf{x})\in I(\mathbf{O}\cup\mathbf{S}) \\ 0.5 & \text{otherwise.} \end{cases}$$

- Lookup-tables enable efficient evaluation over large search regions.

Distractor-aware Object Model

- Identify **visually distracting regions D** whenever they appear within the field-of-view and suppress them in advance

$$P(\mathbf{x}\in\mathcal{O}\,|\,\mathbf{O},\mathbf{D},b_{\mathbf{x}}) = \begin{cases} \frac{H^I_{\mathbf{O}}(b_{\mathbf{x}})}{H^I_{\mathbf{O}}(b_{\mathbf{x}})+H^I_{\mathbf{D}}(b_{\mathbf{x}})} & \text{if } I(\mathbf{x})\in I(\mathbf{O}\cup\mathbf{D}) \\ 0.5 & \text{otherwise.} \end{cases}$$



- Combine both object models with weighting parameter λ

$$P(\mathbf{x}\in\mathcal{O}\,|\,b_{\mathbf{x}}) = \lambda P(\mathbf{x}\in\mathcal{O}\,|\,\mathbf{O},\mathbf{D},b_{\mathbf{x}}) + (1-\lambda)P(\mathbf{x}\in\mathcal{O}\,|\,\mathbf{O},\mathbf{S},b_{\mathbf{x}}).$$

- Regularly update model to handle changing object appearance using learning rate η

$$P_{1:t}(\mathbf{x}\in\mathcal{O}\,|\,b_{\mathbf{x}}) = \eta P(\mathbf{x}\in\mathcal{O}\,|\,b_{\mathbf{x}}) + (1-\eta)P_{1:t-1}(\mathbf{x}\in\mathcal{O}\,|\,b_{\mathbf{x}}).$$

Localization

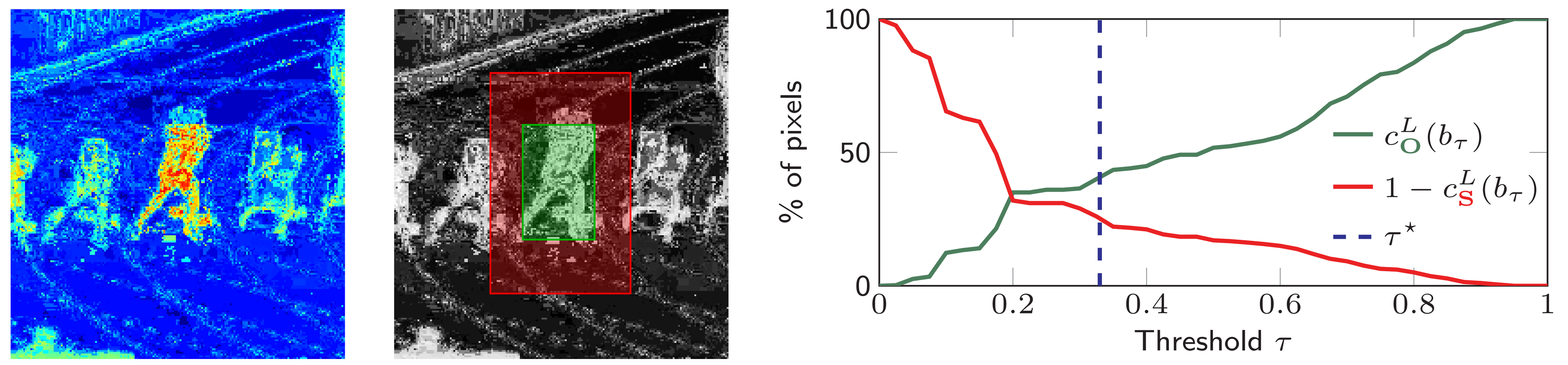
- We follow the widely used **tracking-by-detection** principle.
- Densely sample hypotheses $\mathbf{O}_{t,i}$ within rectangular search region and compute their vote score s_v and distance score s_d

$$s_v(\mathbf{O}_{t,i}) = \sum_{\mathbf{x}\in\mathbf{O}_{t,i}} P_{1:t-1}(\mathbf{x}\in\mathcal{O}\,|\,b_{\mathbf{x}}), \qquad s_d(\mathbf{O}_{t,i}) = \sum_{\mathbf{x}\in\mathbf{O}_{t,i}} \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}_{t-1}\|^2}{2\sigma^2}\right).$$

- We perform an iterative non-maximum suppression to obtain both the new target location $\mathbf{O}^\star_t = \arg\max_{\mathbf{O}_{t,i}}(s_v(\mathbf{O}_{t,i})s_d(\mathbf{O}_{t,i}))$ and potential distractors (high vote score).

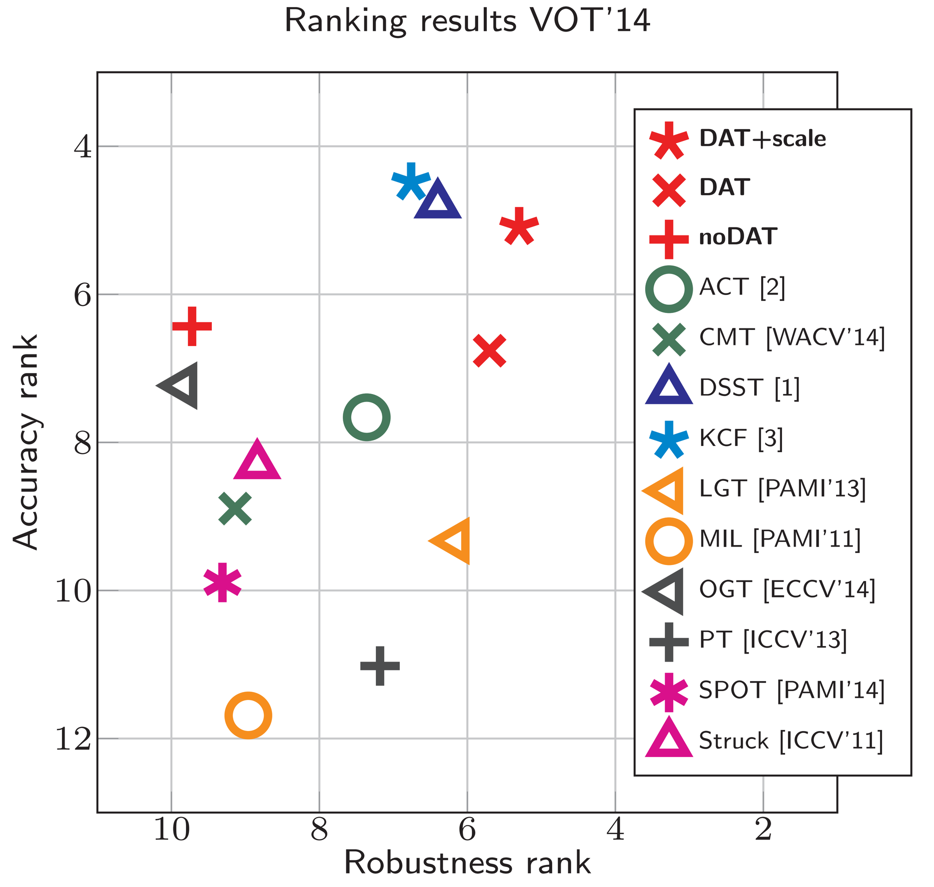
Scale Adaptation

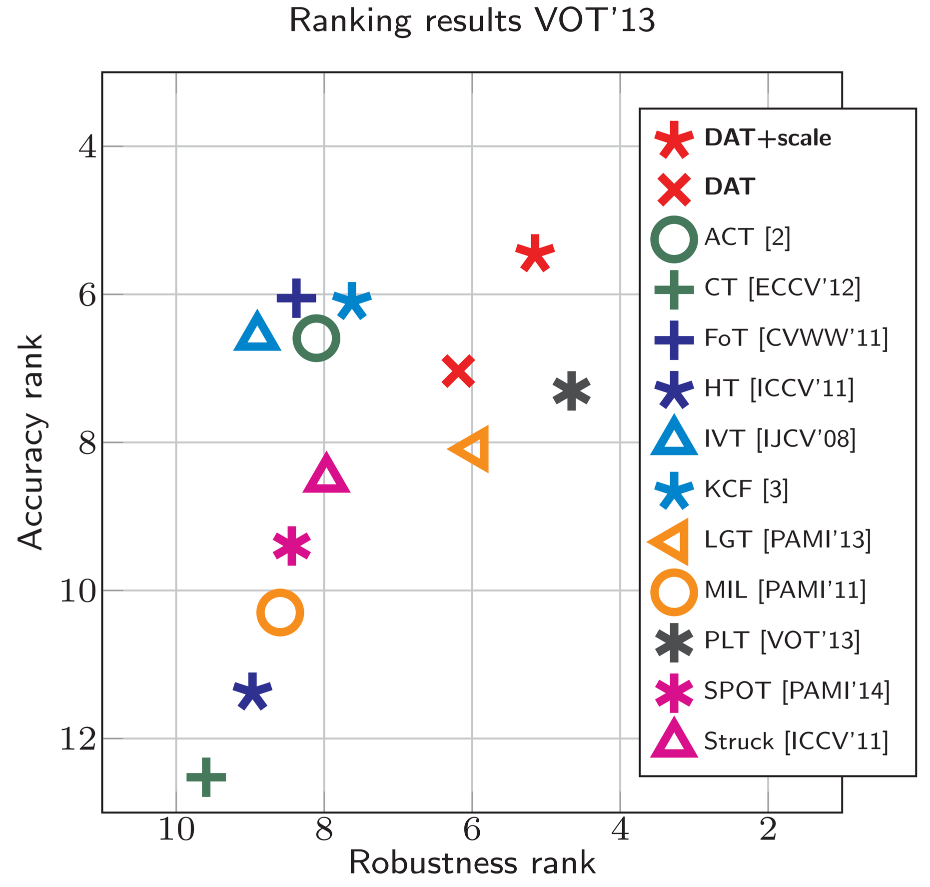
- Segment the object using an **adaptive threshold** $\tau^\star$ based on cumulative histograms $c^L_{\{\mathbf{O},\mathbf{S}\}}$ over the likelihood map L.
- Perform connected component analysis to adapt the target scale.



Benchmark Results

- Extensive evaluations on the **Visual Object Tracking (VOT)** benchmarks [4] show state-of-the-art accuracy and improved robustness.
- We demonstrate **benefits of distractor-awareness** (DAT) and scale-adaptation (DAT+scale) compared to baseline (noDAT) and state-of-the-art trackers (including the challenge winners, *i.e.* DSST & PLT).
- Ranking plots based on statistical significance of performance differences *w.r.t.* accuracy and robustness metrics (Top-performing trackers are located top-right):





- Robustness to noisy initializations (Best, second, and third best results are highlighted):

Tracker	Accuracy		Robustness		Combined
	Score↑	Rank↓	Score↓	Rank↓	Rank↓
ACT [2]	0.49	5.02	1.77	4.56	4.79
DSST [1]	0.57	3.10	1.28	3.98	3.54
KCF [3]	0.57	3.44	1.51	4.28	3.86
LGT [PAMI'13]	0.46	5.12	0.64	3.54	4.33
Struck [ICCV'11]	0.48	5.42	2.22	5.00	5.21
DAT	0.55	3.20	1.06	3.38	3.29
DAT+scale	0.58	2.70	1.03	3.26	2.98

References and Acknowledgments

[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*, 37(3):583–596, 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.

This work was supported by the Austrian Science Foundation (FWF) project Advanced Learning for Tracking and Detection in Medical Workflow Analysis (I535-N23).

Both authors contributed equally.