Kyushu University
Japan

[IS3-2-28]
How to Aggregate Arbitrary Online Trackers

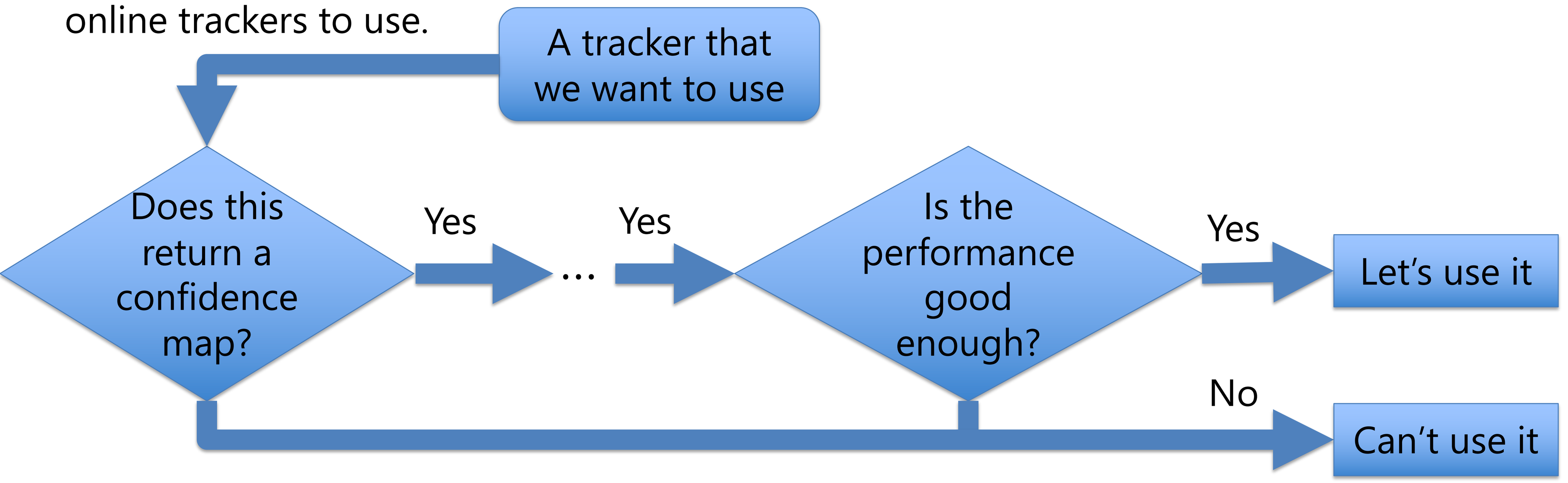
Heon Song, Daiki Suehiro, Seiichi Uchida

The 23rd Meeting on Image Recognition and Understanding

MIRU 2020

Summary

- Visual object tracking is a task that predicts a target location in an image sequence given the initial frame image and target location.
- In object tracking, it is difficult to realize an accurate and stable tracker due to large variations in the target appearance.
- Using multiple trackers can stabilize tracking, but it's difficult to use any trackers we want because most algorithms using multiple trackers require a lot of restrictions to hire online trackers to use.



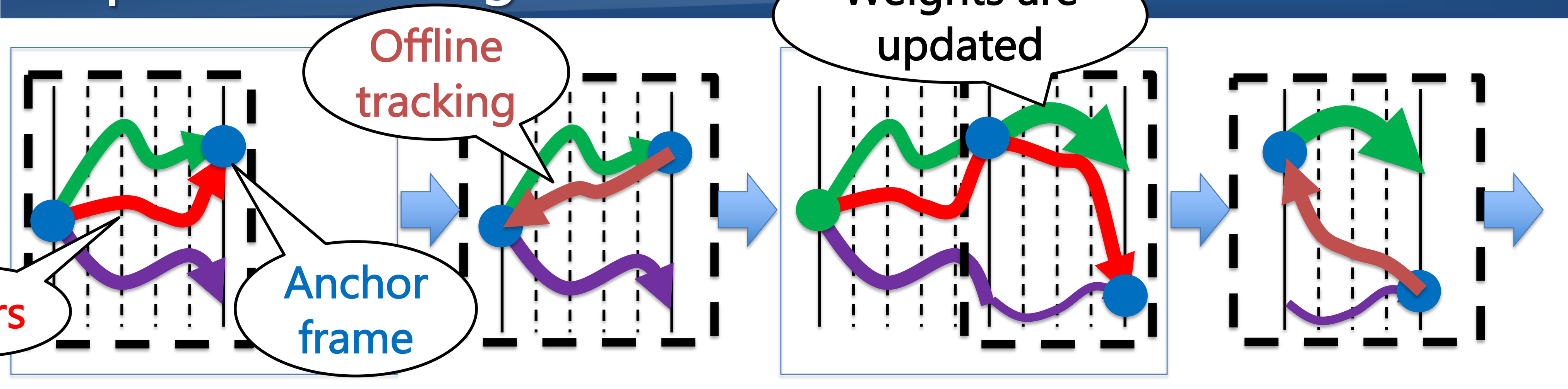
- We propose tracking method that aggregates arbitrary trackers based on expert aggregation technique[2] with feedback.
- Assuming there are some anchor frames where we can specify the target location, the target locations between the two anchor frames in the past can be estimated with high accuracy and used as feedback by offline tracker.

Frame 1 … Frame t … Frame t + 1 (anchor frame)

- The proposed method is theoretically guaranteed to follow the best tracker without any assumption.

[1] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In Proc. CVPR,2008.
[2] K. Quanrud and D. Khashabi. Online learning with adversarial delays. In Proc. NIPS, 2015

Proposed Tracking Method



Aggregate based on initial weight | Evaluate trackers with offline tracker | Aggregate based on updated weight | Evaluate trackers with offline tracker

Step1. Prediction

- (1) At each frame t, N online trackers estimate the target locations.
- (2) Predict the current target location by selecting one of the estimations by using the trackers' weight as a probability distribution
- (3) If we have a reliable target location, determine frame t as an anchor frame.

Step2. Feedback – only at anchor frame

- (3) The feedback is given by the offline tracker[1] between previous anchor frame u and current anchor frame t.

Step3. Evaluation – only at anchor frame

- (4) Calculate the cumulative tracking error of ith tracker $L^{u+1:t}$ between previous anchor frame u and current frame t and update weights so that trackers that are likely to be the best tracker have a higher weight.

$$w_i^{t+1} = \frac{w_i^t \exp\left(-\eta L_i^{u+1:t}\right)}{\sum_{j=1}^{N} w_j^t \exp\left(-\eta L_j^{u+1:t}\right)}$$

Theoretical Guarantee[2]

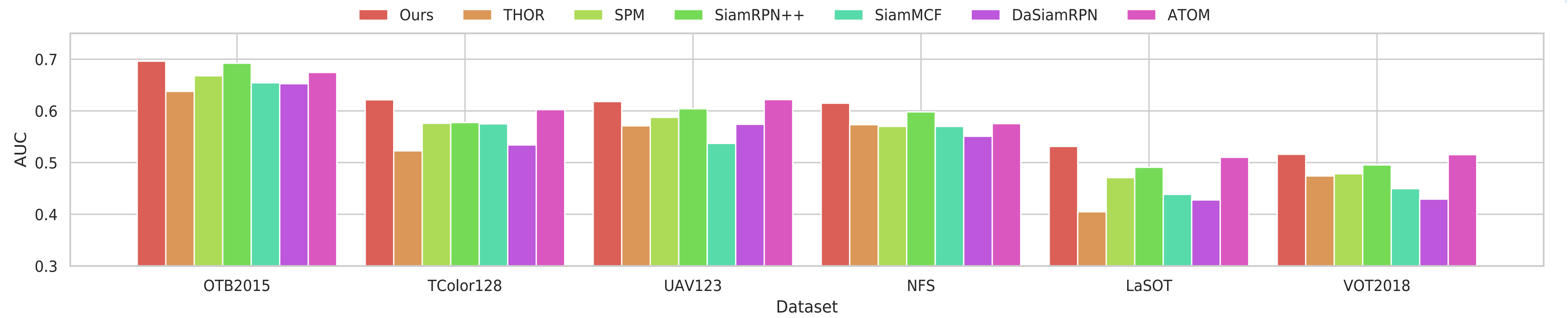
- By updating the weights as above, the regret is upper bounded, which means that we can theoretically guarantee ours can follow the best tracker even in the worst case. It doesn't matter which trackers you aggregate for any videos.

$$Regret = \underbrace{L_{ours}^{1:T}}_{\text{Tracking error of ours}} - \underbrace{\min_{i=1,\dots,N} L_i^{1:T}}_{\text{Tracking error of the best}} = O(\sqrt{T \ln N})$$

Experimental Results

- The proposed method and the six state-of-the-art trackers are evaluated with average area-under-curve score(AUC) of overlap.

Performance of trackers



Ranking of trackers by AUC

- No tracker was always the best, and the best expert drastically changed over the videos.
- Ours achieved the second or third best performance for most videos.

Tracking examples