Staple: Complementary Learners for Real-Time Tracking

Luca Bertinetto, Jack Valmadre, Stuart Golodetz, Ondrej Miksik, Philip Torr

Torr Vision Group, Department of Engineering Science, University of Oxford, UK

CVPR 2016

UNIVERSITY OF OXFORD

Code available!

github.com/bertinetto/staple

Model-free, short-term, single-target tracking

Problem: track an arbitrary object selected online.

- Sole supervision: bounding box at first frame.
- *Model-free*: agnostic to the object's class.
- *Short-term*: no re-detection logic.

Main challenges

- Large variability of target objects and video scenarios.
- *Stability-Plasticity* dilemma.

Related work

Correlation filters

- Dense sampling of object and background.
- Fast evaluation: sampling space is a circulant matrix and can be diagonalized with DFT.

from [Henriques12]

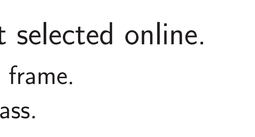
Colour histograms

- Dense sampling of object and background.
- Fast evaluation with Integral Images.

from [Possegger15]

Deep trackers

- Benefit from expressiveness of conv-net features.
- Require to perform fine-tuning online: slow (far from real-time).

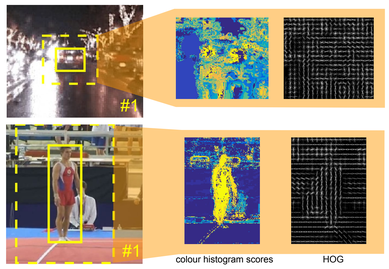


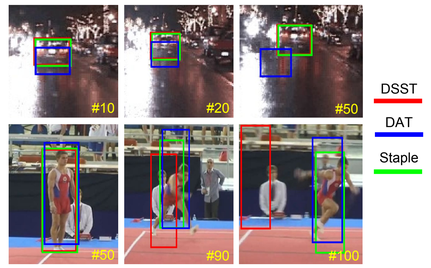
from [Nam15]

Ensemble methods

- Run many trackers in parallel, modelling individual confidences.
- Merge final estimations only. (*e.g.* [Kwon11], [Wang14], [Zhanh14])

Motivation





- Colour distributions rely on pixel values to discriminate target from background→no concept of locality: robust to shape changes, sensitive to blur and poor illumination.
- Template models rely on spatial configuration →robust to blur and poor illumination, sensitive to shape changes.

Formulation

Staple = **S**um of **T**emplate **a**nd **P**ixel-wise **Le**arners.

- Combination of score functions f_{tmpl} and f_{hist} evaluated on complementary cues

$$f(x) = \gamma_{\text{tmpl}} f_{\text{tmpl}}(x) + \gamma_{\text{hist}} f_{\text{hist}}(x)$$

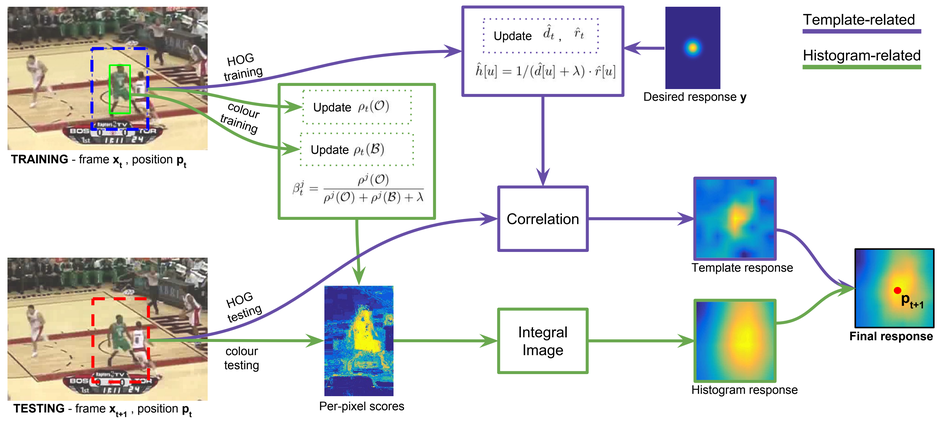
- $f_{\text{tmpl}}(x; h)$: linear function of HOG feature image ϕ_x.
- $f_{\text{hist}}(x; \beta)$: average of a score image computed from the histogram model β.
- Two ridge regression problems at each t depending on previous images/locations $\mathcal{X}_t$:

$$h_t = \arg\min_h \left\{ L_{\text{tmpl}}(h; \mathcal{X}_t) + \tfrac{1}{2}\lambda_{\text{tmpl}} \|h\|^2 \right\}$$

$$\beta_t = \arg\min_\beta \left\{ L_{\text{hist}}(\beta; \mathcal{X}_t) + \tfrac{1}{2}\lambda_{\text{hist}} \|\beta\|^2 \right\}$$

- Correlation (in Fourier domain) and Integral Image→dense responses with fast sliding window search.

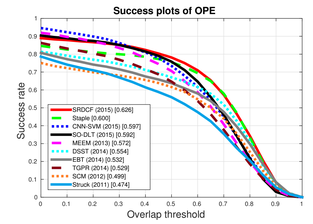
Pipeline

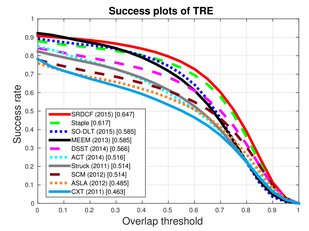


Results

Despite its extreme simplicity, **Staple outperforms the state-of-the-art** on multiple benchmarks, yet running at **80 fps**.

OTB-13

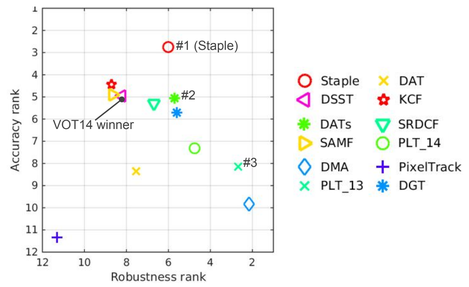




VOT 2014

Tracker	Year	Where	Accuracy	# Failures	Overall Rank	Speed (fps)
Staple	2016	CVPR	0.644	9.38	4.37	80
DATs	2015	CVPR	0.580	13.17	5.39	15
PLT_13	2013	VOT	0.523	1.66	5.41	76
DGT	2014	TIP	0.534	13.78	5.66	10
SRDCF	2015	ICCV	0.600	15.90	5.99	5
DMA	2015	CVPR	0.476	0.72	6.00	8
PLT_14	2014	VOT	0.537	3.41	6.03	63
KCF	2015	PAMI	0.613	19.79	6.58	80
DSST	2014	BMVC	0.607	16.90	6.59	24
SAMF	2014	ECCVw	0.603	19.23	6.79	7
DAT	2015	CVPR	0.519	15.87	7.95	17
PixelTrack	2013	ICCV	0.420	22.58	11.31	**114**

Table 1 : Recent trackers on the VOT14 benchmark.



VOT 2015

Tracker	Year	Where	Accuracy	# Failures	Rank	Speed (fps)
MDNet	2015	ICCV	0.583	0.69	14.31	1
DeepSRDCF	2015	VOT	0.528	1.05	19.16	< 1
SRDCF	2015	ICCV	0.521	1.24	21.01	5
Staple	2016	CVPR	0.533	1.39	21.64	**80**
SO-DLT	2015	arXiv	0.535	1.78	22.71	5
NSAMF	2015	VOT	0.490	1.29	22.93	5
EBT	2015	arXiv	0.453	1.02	23.01	5
sPST	2015	ICCV	0.508	1.48	23.04	2
RAJSSC	2015	VOT	0.518	1.63	23.53	2
SC-EBT	2015	ICML	0.523	1.86	23.70	-

Table 2 : VOT15 top 10 (of 63)

Supported by Apical - www.apical.co.uk

luca@robots.ox.ac.uk

www.robots.ox.ac.uk/~luca/staple.html