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

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

Problem: track any given object in a video.

- *Model-free* - algorithm has to be agnostic on the class of the object.
- *Short-term* - full occlusions not addressed, *i.e.* no re-detection logic.
- *Single-target* - no data association problem.

Main challenges

- Huge variety of scenes and transformations.
- *Stability-Plasticity* dilemma.

Related work and motivation

Correlation filters

- Fast evaluation in Fourier domain.
- Dense sampling of target object and surroundings.
- HOG: robust to blur, illumination changes, sensitive to deformation.

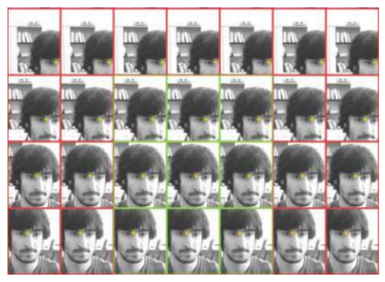


Figure 1 : Sampling space is a circulant matrix and can be diagonlized with DFT (image from [Henriques12]).

Colour histograms

- Fast evaluation with Integral Images.
- Dense sampling.
- Colour statistics, no concept of locality: sensitive to blur, illumination changes, robust to deformation.

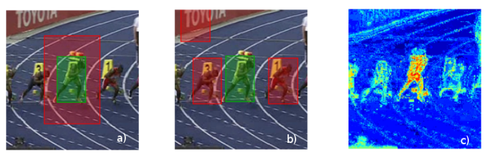


Figure 2 : a) Samples for object and surrounding. b) Samples for object and distractors. c) Likelihood map obtained combining models built on a) and b) (image from [Possegger15]).

Ensemble methods

- Run several trackers in parallel to alleviate their inaccuracies.
- Model individual trackers' confidence.
- Merge final estimations only.

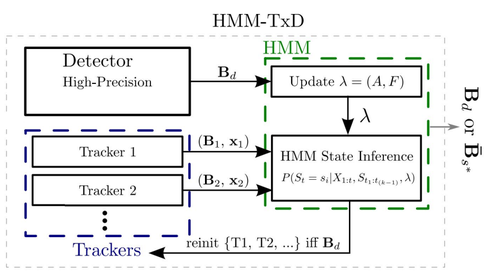


Figure 3 : HMM-TxD merges the final prediction of an array of independent trackers and a detector (image from [Vojir15]).

Formulation

- Staple: **S**um of **T**emplate **a**nd **P**ixel-wise **Le**arners, *i.e.* combined score function $f(x) = \gamma_{\text{tmpl}} f_{\text{tmpl}}(x) + \gamma_{\text{hist}} f_{\text{hist}}(x)$ from:
  - Complementary cues
  - Compatible (dense) responses
- Compatibility of responses assured by objective functions, both with target $[0 - 1]$ in ridge regression framework.
  - $f_{\text{tmpl}}(x; h) = \sum_{u \in \mathcal{T}} h[u]^T \phi_x[u]$.
  - $f_{\text{hist}}(x; \beta) = g(\psi_x; \beta) = \frac{1}{|\mathcal{H}|} \sum_{u \in \mathcal{H}} \zeta_{(\beta,\psi)}[u]$.
- Both template and histogram features are *feature images*; we can use correlation in Fourier and Integral Image for fast sliding window search.

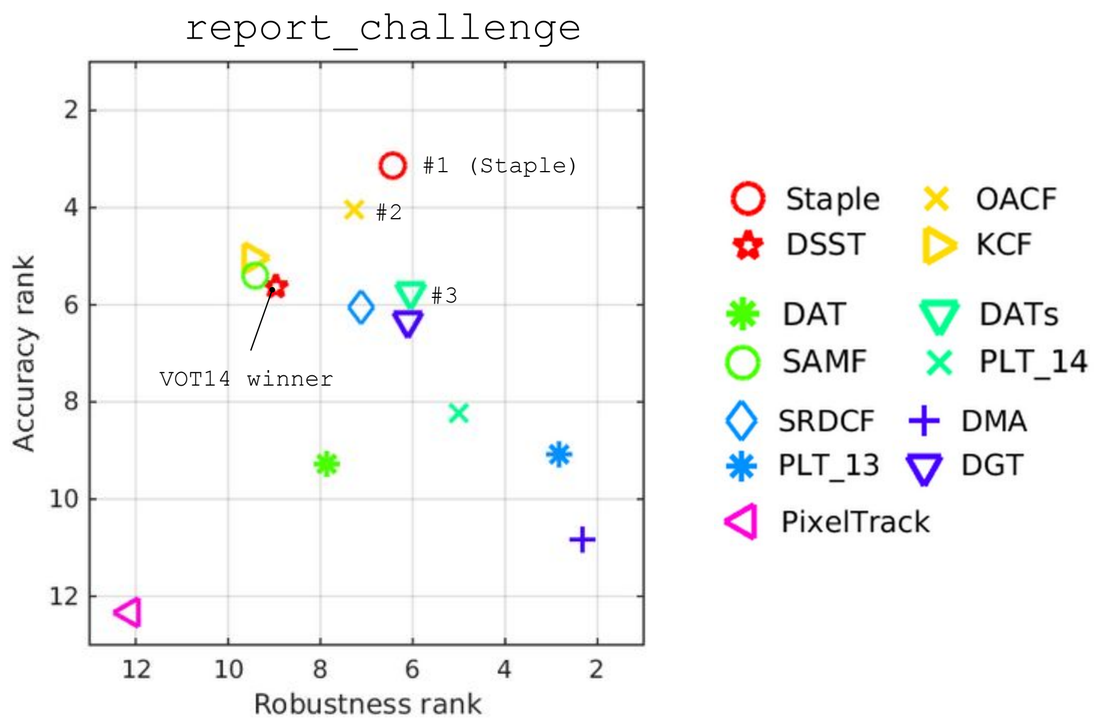
Results

Despite its simplicity, **Staple outperforms the state-of-the-art** and runs at approximately **90 FPS** (on a 4GHz machine).

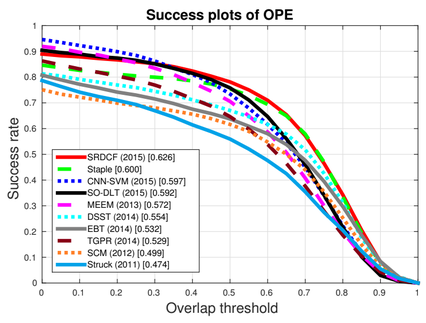
- No dataset overfitting: VOT15 used as validation fold, VOT14 and OTB-13 as test folds.
- Up-to-date comparison with very recent trackers.

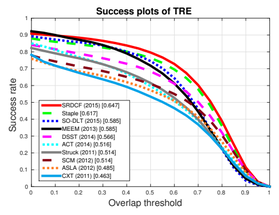
VOT14

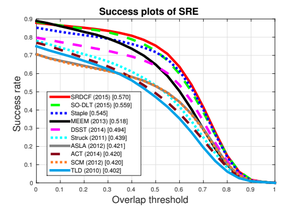
Tracker	Year	Where	Accuracy	# Failures	Overall Rank
Staple	-	-	0.644	9.38	4.78
OACF	2015	VOT2015	0.621	15.56	5.66
DATs [33]	2015	CVPR	0.580	13.17	5.89
PLT_13 [24]	2013	VOT13	0.523	1.66	5.95
DGT [6]	2014	TIP	0.534	13.78	6.22
DMA [40]	2015	CVPR	0.476	0.72	6.58
SRDCF [9]	2015	ICCV	0.600	15.90	6.59
PLT_14 [24]	2014	VOT14	0.537	3.41	6.61
KCF [18]	2015	PAMI	0.613	19.79	7.25
DSST [8]	2014	BMVC	0.607	16.90	7.30
SAMF [29]	2014	ECCVw	0.603	19.23	7.41
DAT [33]	2015	CVPR	0.519	15.87	8.58
PixelTrack [11]	2013	ICCV	0.420	22.58	12.13



OTB-13







Staple pipeline