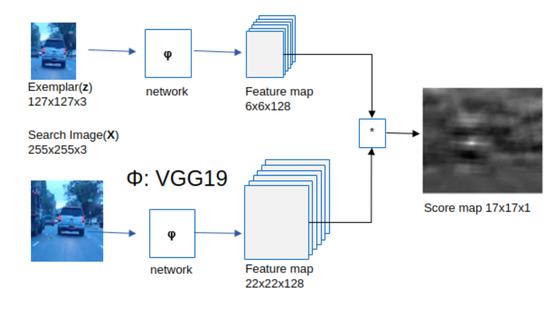
SIAMESE FULLY CONVOLUTIONAL TRACKER WITH MOTION CORRECTION

Mathew Francis, Prithwijit Guha
{m.francis,pguha}@iitg.ac.in
Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati

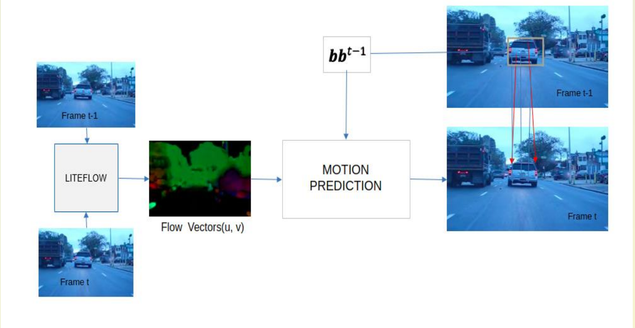
Abstract

Visual tracking algorithms use cues like appearance, structure, motion etc. for locating an object in a video. We propose an ensemble tracker with two components. First, a Siamese tracker that learns object appearance from a static image. Second, motion information obtained from consecutive frames using a flow estimation network. The motion information is used to correct the predictions obtained by the appearance-based tracking component of the ensemble. Complementary nature of the two components (appearance and motion) lead to performance improvement as observed in

Appearance Component



Motion Component



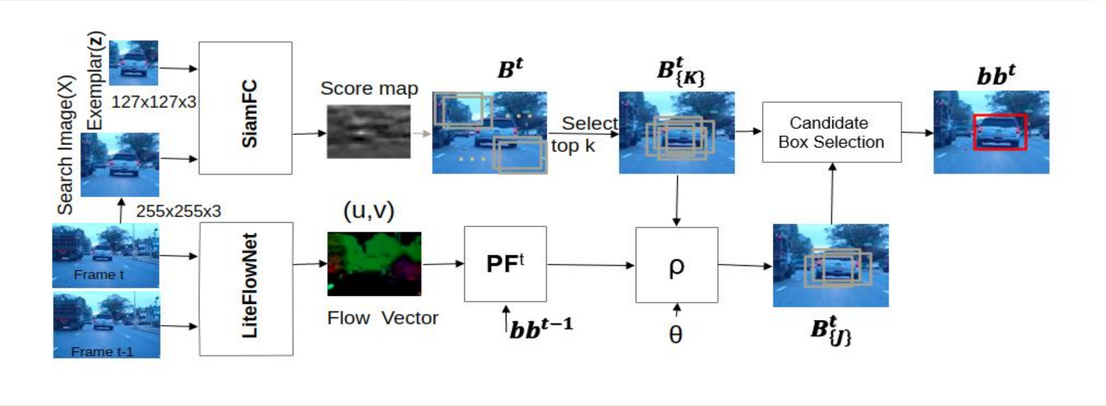
VOT2019

Tracker	Overlap	Failures	EAO
NCC	0.4636	173.0405	0.0846
SiamFC-MC	0.5093	53.6017	0.1845
SiamFCOSP	0.5008	71.7348	0.1707
SiamMask	0.5907	29.3580	0.2856
SiamRPNX	0.5190	35.7638	0.2237
RSiamFC	0.4659	58.7354	0.1627
RankingT	0.5241	22.5801	0.2698
DPT	0.4832	60.1490	0.1587
FSC2F	0.4711	48.5677	0.1850
CSRDCF	0.4916	40.7859	0.2014
KCF	0.4377	78.3864	0.1135
MIL	0.3884	78.6186	0.1201
Struck	0.4120	100.9718	0.0963
SSRCCOT	0.4895	33.2756	0.2330
ASMS	0.4731	55.2126	0.1595

VOT2018

Tracker	Overlap	Failures	EAO
MBSiam	0.5242	26.2737	0.2411
Dsiam	0.5089	40.0874	0.1963
SiamFC	0.5002	34.0259	0.188
SiamFC-MC	0.5121	36.9841	0.2013
SiamRPN	0.5779	17.6608	0.3827
SiamVGG	0.5254	20.4526	0.2865
SRDCF	0.4802	64.1136	0.1189
Staple	0.5244	44.0194	0.1694
UpdateNet	0.509	26.8721	0.2436
ASMS	0.4884	36.5313	0.1692
CSRDCF	0.4846	23.5731	0.2561
DCFNet	0.4647	35.2015	0.1825
DeepCSRDCF	0.4827	19.0067	0.2926
DSST	0.3895	95.5587	0.0788
ECO	0.4758	17.6628	0.2805
FoT	0.3901	61.5017	0.1299
KCF	0.444	50.0994	0.1349
L1APG	0.4209	129.5924	0.0693
MIL	0.3847	64.3029	0.1183
RAnet	0.4419	47.4719	0.1415

Proposed Approach



Problem Formulation

$$s_i = f(Z, x^i) = \varphi(Z) * \varphi(x^i) + b, b \in R$$

$$\rho_i^t = \frac{|\Gamma \cap \Gamma^i|}{|\Gamma^i|}$$

$*$: *correlation*
ρ: *combined score*

Γ: *set of object pixel locations as predicted by the motion component*

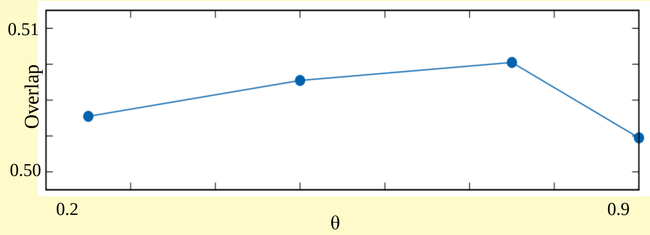
Γ^i: *set of object pixel locations as predicted by the appearance component*

$$J^t = \{i | \rho_i^t > \theta; i \in K^t\}$$

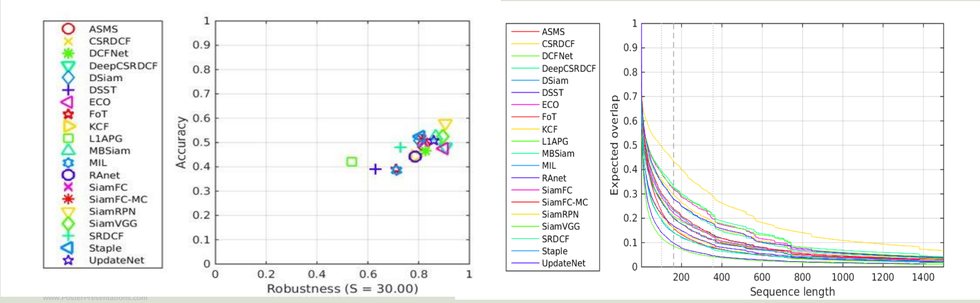
K^t: *set of indices corresponding to top K scores in S*

$$bb^t := bb_j^t, \qquad j = \begin{cases} \operatorname{argmax}_{i \in K^t}(s_i^t), & |J^t| = 0 \\ \operatorname{argmax}_{i \in J^t}(\rho_i^t), & Otherwise \end{cases}$$

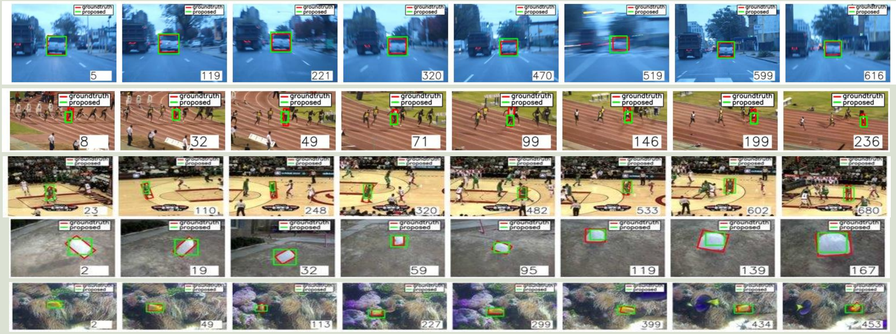
Effect of Parameter



Performance



Visual Output



Notable References

- Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A. and Torr, P.H., 2016, October. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision* (pp. 850-865). Springer, Cham.
- Hui, T.W., Tang, X. and Change Loy, C., 2018. Liteflownet: A lightweight convolutional neural network for optical flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8981-8989).
- Tao, R., Gavves, E. and Smeulders, A.W., 2016. Siamese instance search for tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1420-1429).