Real-time tracker with fast recovery from target loss

Alessandro Bay
Cortexica Vision Systems, London, UK

Eduard Vazquez
AnyVision, Belfast, UK

Panagiotis Sidiropoulos
Space and Climate Physics Department, UCL, London, UK

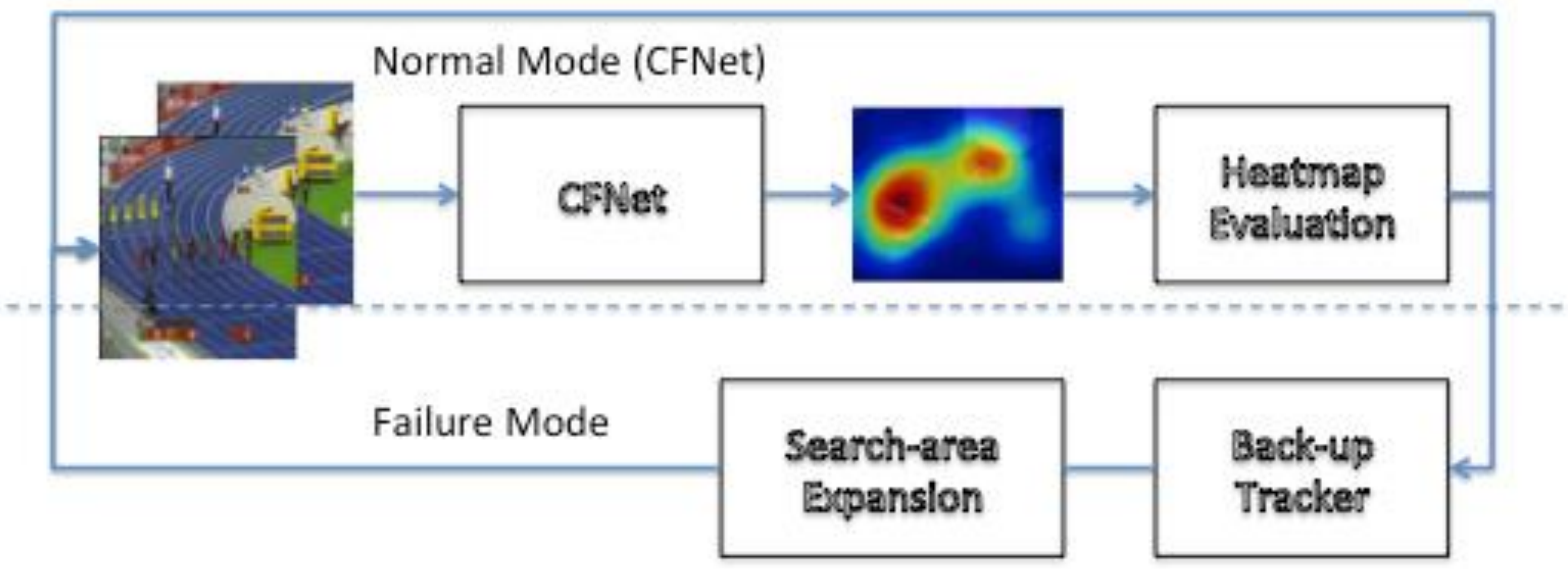
Michele Sasdelli
Australian Institute for Machine Learning, Adelaide, Australia

Abstract

In this paper, we introduce a variation of a state-of-the-art real-time tracker (CFNet), which adds to the original algorithm robustness to target loss without a significant computational overhead. The new method is based on the assumption that the feature map can be used to estimate the tracking confidence more accurately. When the confidence is low, we avoid updating the object's position through the feature map; instead, the tracker passes to a single-frame failure mode, during which the patch's low-level visual content is used to swiftly update the object's position, before recovering from the target loss in the next frame. The experimental evidence provided by evaluating the method on several tracking datasets validates both the theoretical assumption that the feature map is associated to tracking confidence, and that the proposed implementation can achieve target recovery in multiple scenarios, without compromising the real-time performance.

CFNet – Fast Target Loss Recovery (FTLR)

- We address the *sampling drift* issue in the tracking-by-similarity CFNet
- The localization ambiguity is detected using the Nearest Neighbour Distance Ratio (NNDR)
- If localisation is declared ambiguous, the tracker enters a *failure mode*
  - the tracking is conducted through low-level image content modelling
  - the search area is expanded
- Besides, an improved running average method is involved to update the query model from the previously seen feature maps
- Improved accuracy, without increasing the computational time



Failure mode

- The object position is updated according to the estimation of one the following backup trackers
  1) No update until the NNDR is above the confidence threshold
  2) Bilinear interpolation using the past two frames
  3) Census transform (Hirshmuller and Scharstein) correlation on image patches
     - Linear computational complexity
     - Object edges are preserved
     - Robust to radiometric differences
     - Robust optical flow estimation
  4) Ground truth position (theoretical upper bound for accuracy)
- In the following frame the object is searched in a wider area

Smooth running average (SA)

- The strategy used in CFNet to create the query model is a simple running average

$$\begin{cases} Q_1 = F_1 \\ Q_n = (1-\alpha)Q_{n-1} + \alpha F_n \end{cases}$$

where Q_n is the n-th query model and F_n is the n-th feature map

- The dependency from the first frames is reduced as follows

$$\begin{cases} Q_1 = F_1 \\ Q_n = (1-\frac{0.5}{n}-\alpha)Q_{n-1} + (\frac{0.5}{n}+\alpha)F_n \end{cases}$$

Results on OTB dataset (Wu and Yang)

		OTB-2013				OTB-50				OTB-100			
		OPE		TRE		OPE		TRE		OPE		TRE	
Method	fps	IoU	prec	IoU	prec	IoU	prec	IoU	prec	IoU	prec	IoU	prec
CFNet (Valmadre et al.)	71.1	57.0	74.1	59.9	76.1	49.4	64.3	52.9	69.0	55.7	71.6	58.2	73.7
CFNet-FTLR_0	64.2	57.4	74.4	60.2	76.5	49.1	64.1	53.0	69.2	54.6	70.3	58.0	73.4
CFNet-FTLR_1	64.0	57.9	75.2	60.2	76.5	48.9	63.8	53.0	69.0	54.8	69.9	58.1	73.6
CFNet-FTLR	61.6	**60.0**	**78.2**	60.6	76.9	51.3	68.1	53.9	70.7	**57.3**	74.1	58.7	74.4
CFNet-FTLR_SA	62.3	58.7	76.8	**61.9**	**79.8**	**51.5**	**68.5**	**55.0**	**73.2**	57.1	**74.6**	**60.3**	**77.8**
CFNet-FTLR_GT	65.0	62.7	83.6	65.3	84.6	54.6	74.6	58.7	79.3	59.3	78.2	62.5	80.8

Results on UAV-123 dataset (Mueller et al.)

	OPE		TRE	
Method	IoU	prec	IoU	prec
KCF (Henriques et al.)	33.1	52.3	–	–
DSST (Danelljan et al.)	35.6	58.6	–	–
CFNet (Valmadre et al.)	47.0	66.5	52.4	72.2
CFNet-FTLR	47.2	67.0	**53.3**	73.7
CFNet-FTLR_SA	**47.6**	**67.6**	52.9	**73.8**
CFNet-FTLR_GT	54.9	80.7	59.2	84.5

Results on DTB-70 dataset (Li and Yeung)

	OPE		TRE	
Method	IoU	prec	IoU	prec
DSST (Danelljan et al.)	26.4	40.2	–	–
KCF (Henriques et al.)	28.0	46.8	–	–
CFNet (Valmadre et al.)	39.4	57.9	48.1	67.2
CFNet-FTLR	41.2	61.3	49.1	68.7
CFNet-FTLR_SA	**43.0**	**64.5**	**50.7**	**72.2**
CFNet-FTLR_GT	52.4	82.2	56.3	81.9

[1] M. Danelljan, A. Robinson, F. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking", *ECCV 2016*.

[2] J.F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters", *TPAMI 2015*.

[3] H. Hirschmuller, and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences", *TPAMI 2009*.

[4] S. Li, and D.-Y. Yeung, "Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models", *AAAI 2017*.

[5] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking", *ECCV 2016*.

[6] J. Valmadre, L. Bertinetto, J.F. Henriques, A. Vedaldi, and P.H. Torr, "End-to-end representation learning for correlation filter based tracking", *CVPR 2017*.

[7] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark", *CVPR 2013*.

CORTEXICA
AI SOLUTIONS FOR BUSINESS