ICIP2021 Paper ID: 1161

ROBUST VISUAL OBJECT TRACKING WITH SPATIOTEMPORAL REGULARISATION AND DISCRIMINATIVE OCCLUSION DEFORMATION

Shiyong Lan[1], jin Li[1], Shipeng Sun[1], Xin Lai[2], Wenwu Wang[3]
[1] Sichuan University, China; [2] Southwest Petroleum University, China; [3] University of Surrey, UK

Introduction

Related works

Discriminative Correlation Filters (DCFs) have several advantages:

- Directly learn mapping from dense samples around the target to Gaussian-like soft labels, and not need to manually annotation.
- With the circular shifts of training samples, the convolution operation used in correlation can be efficiently computed in Fourier domain.
- with robust multi-feature representation for target, such as HOG, Color Names (CN) and deep convolutional neural network (CNN) features , the tracking accuracy and robustness of the DCFs can be further improved.
- With spatiotemporal regularisation, the DCFs can outperform both in suppressing the adverse boundary effects and in computation to approximate multiple training samples with single one.

Problem Definition

The performance of DCF methods, such as GFS-DCF, STRCF, HCFstar, degrades significantly when the target is deformed between successive frames, and in this scenario, the response of the x-correlation between the model and the candidate targets will be distorted.

Main Contributions

- We propose a new method based on a spatiotemporal regularised DCF for robust visual tracking, which considers the difference between occlusion-induced deformation and target self-deformation.
- We propose an adaptive strategy for model update according to the difference between occlusion and self-deformation of target. The proposed strategy can not only enhance the tracking performance with rapid target self-deformation, but also reduce tracking drift caused by occlusion.

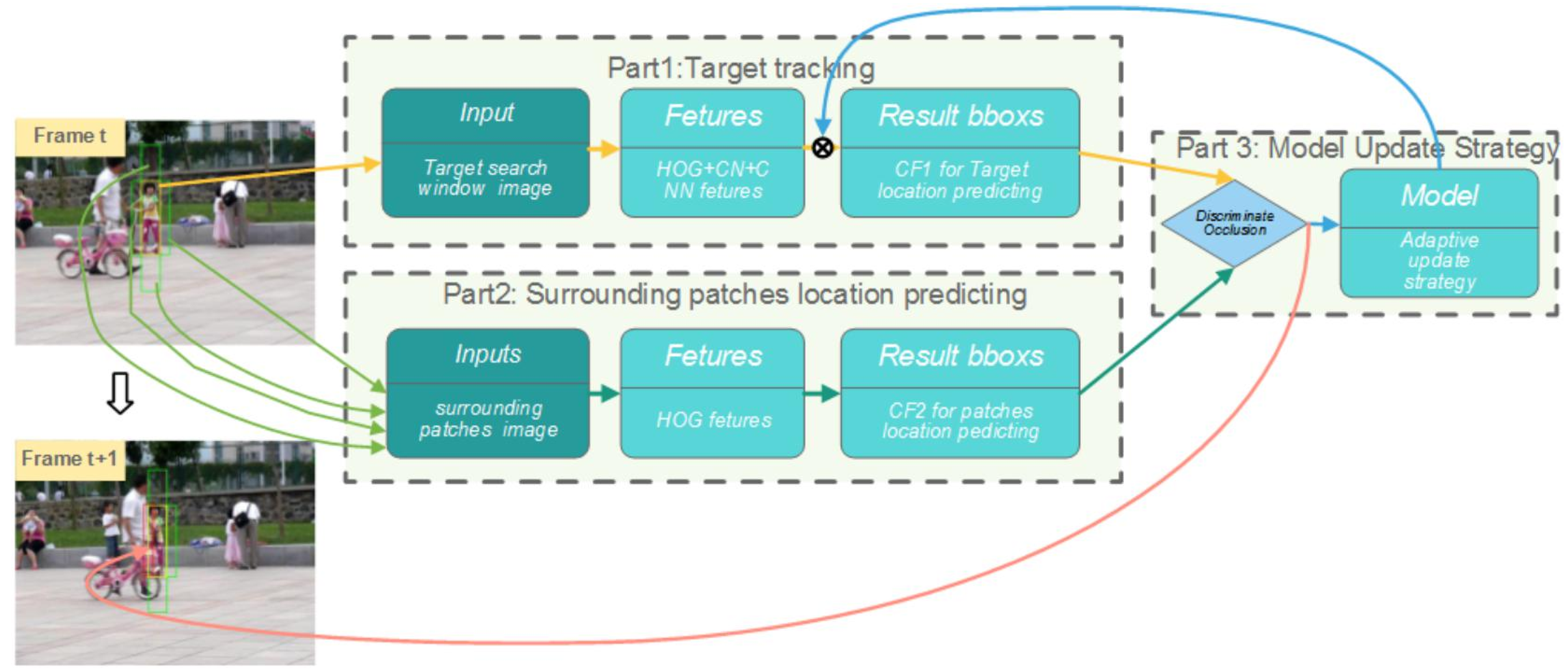
Proposed Method

Motivation

- Different types of deformation of target may require different model update strategies in visual object tracking.
- To equip the target object with its own tactile (i.e. adjacent blocks) and use the tactile sense to discrimate whether the target is occluded or due to self-deformation.

Proposed Method

Architecture



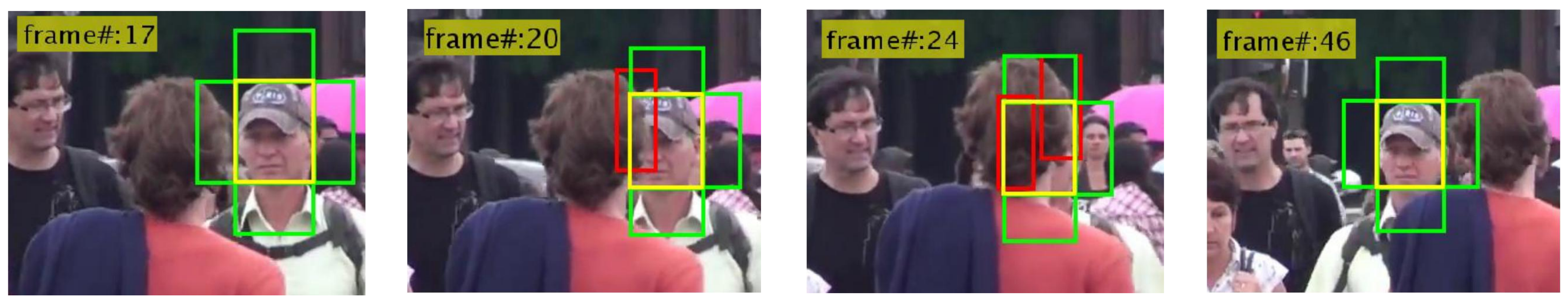
The proposed method consists of three parts: target tracking, prediction of the location of the patches surrounding a target, and an adaptive update strategy based on target deformation information.

Part1. Target tracking

$$\tilde{F}_t = \arg\min_{F} \left\|\sum_{j=1}^{C} F_t^j \circledast X_t^j - Y_t\right\|^2 + \lambda_c \sum_{j=1}^{C}\left\|F_t^j\right\| + \frac{\lambda_1}{2}\sum_{j=1}^{C}\left\|W^j \cdot F_t^j\right\|^2 + \frac{\mu}{2}\left\|F_t - F_{(t-1)}\right\|^2 \quad (1)$$

Where λ_c is the regularization parameter for channel selection, the second term in Eq. (1) focuses on channel sparsity to realize feature channel selection, here, the channels with the lowest channel attributes are set to zero by preset proportion.

Part2. Surrounding patches location predicting



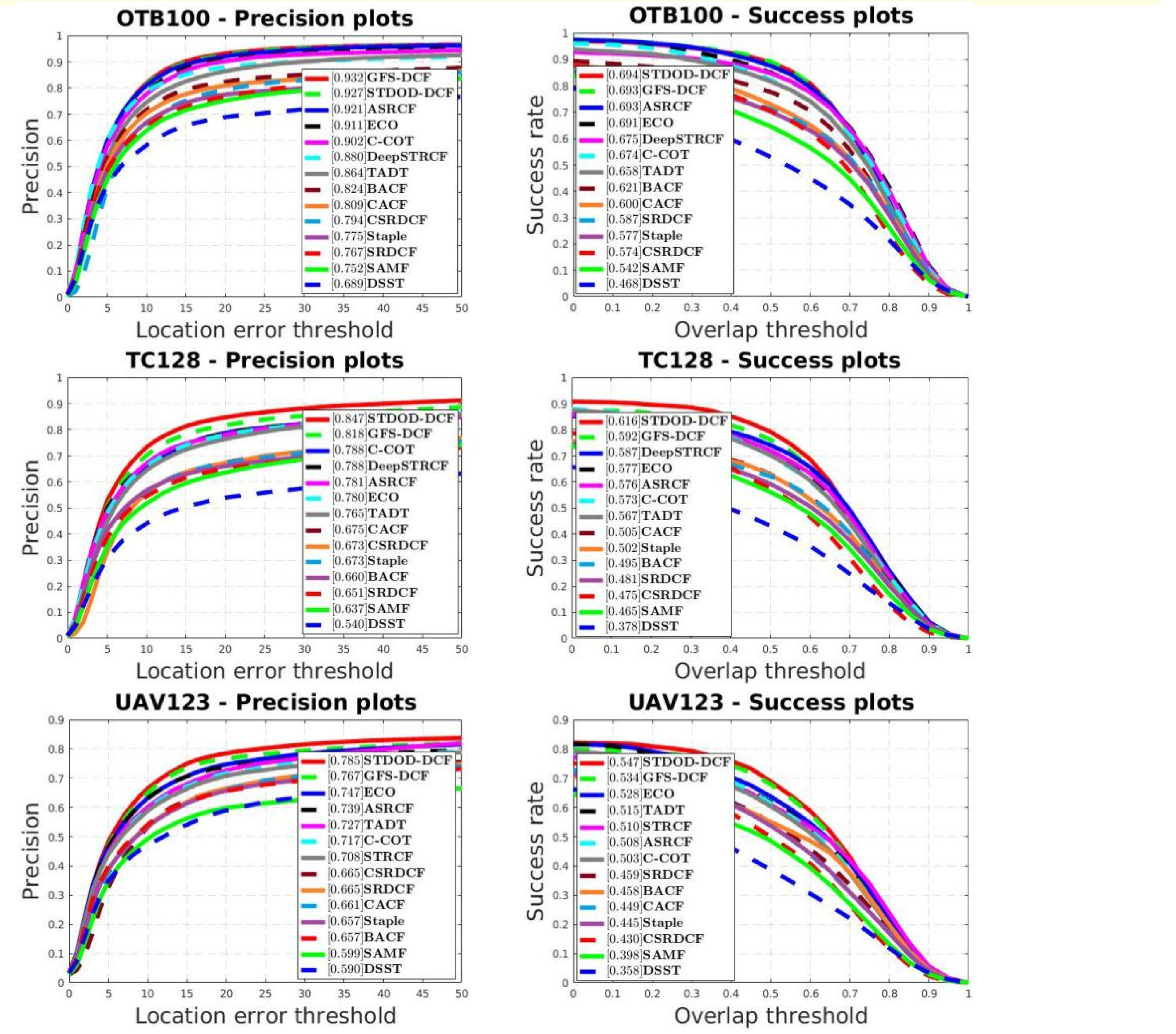
If a patch (red) occludes the target, it will be kept in a set of candidate patches, otherwise it will be removed from the set of patches and reset new one according the target.

Part3. Model updating strategy

The updating rate $\xi \in [0, 1]$ is used to control the update of tracking filter $\mathbf{F}_{t+1}$ as $\xi\mathbf{F}_t + (1-\xi)\mathbf{F}_t$ from the t-th frame, where the ξ will be adjusted according to the discriminative deformation of the target as following:

- If the target is occluded by this surrounding patch, we can down-adjust ξ and limit the model update to avoid the model contamination.
- If no occlusion has been detected, but the confidence level of the x-correlation response mapping from the target is low, we then can up-adjust ξ and enhance the model update to adapt the rapid self-seformation of target.

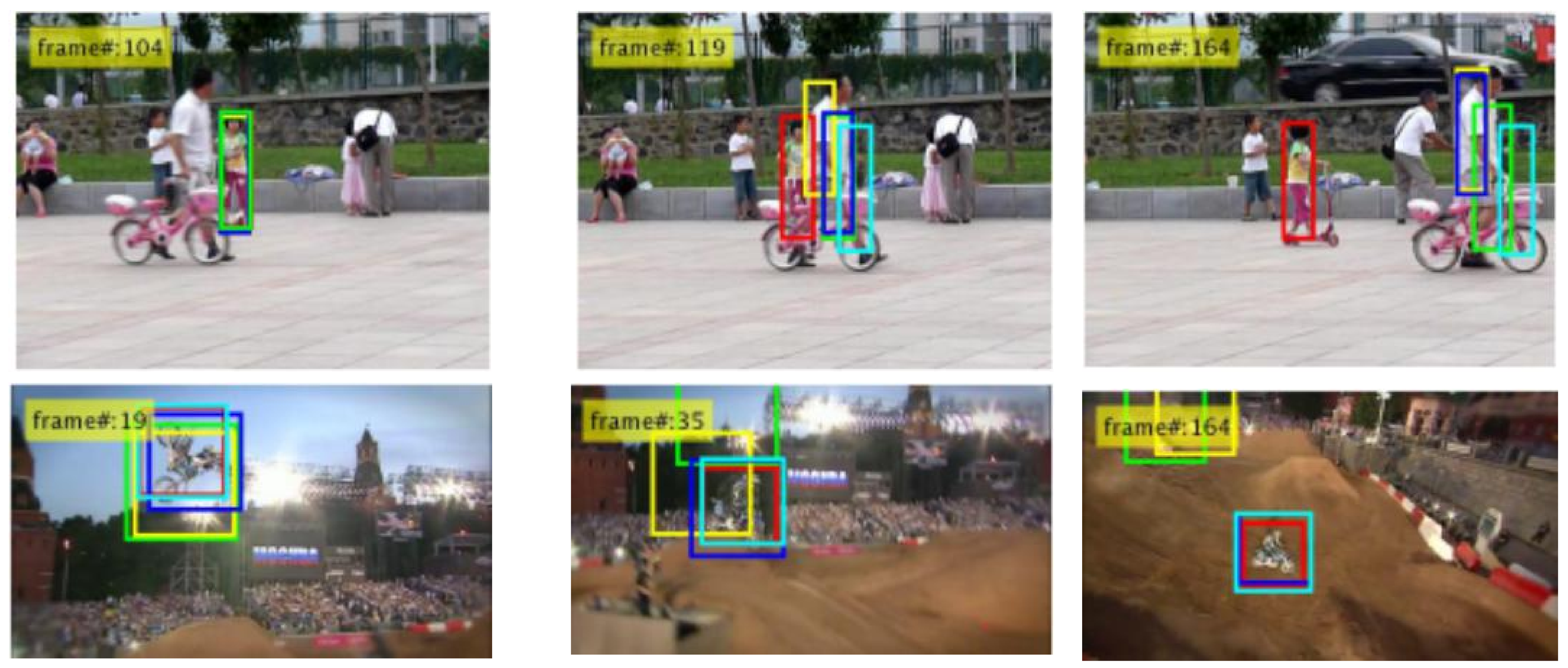
Experimental Results



The precision plots (left) and success plots (right) on OTB100, TC128 and UAV123, where the precision plots show the distance precision (DP) value with a threshold of 20 pixels, and the success plots show the overlap success value with the area under the curve (AUC).

Table 1.VOT2018 unsupervised overlap average overview.
(The red, blue and green represent the best three results respectively.)

	tag_camera_motion	tag_empty	tag_illum_change	tag_motion_change	tag_occlusion	tag_size_change	tag_all
STDOD-DCF	0.5210	0.4136	0.4576	0.4944	0.3560	0.4812	0.4586
GFS-DCF	0.4865	0.3959	0.5007	0.4700	0.3175	0.4578	0.4344
DeepSTRCF	0.4963	0.4074	0.4105	0.4391	0.3225	0.4301	0.4365
LADCF	0.4897	0.3882	0.4089	0.4575	0.3204	0.3885	0.4213
LSART	0.4496	0.4569	0.4316	0.4414	0.2719	0.4016	0.4389
ECO	0.4189	0.4132	0.4251	0.3701	0.2804	0.3695	0.4025
BACF	0.2759	0.2414	0.3254	0.2548	0.2056	0.1759	0.2447
SiamFC	0.3598	0.3489	0.3867	0.3358	0.2385	0.3310	0.3428
KCF	0.2816	0.2431	0.3202	0.2704	0.2628	0.2774	0.2671
MCPF	0.4504	0.4702	0.3606	0.4244	0.2724	0.4560	0.4440
CCOT	0.4002	0.4113	0.3618	0.3386	0.2767	0.3532	0.3909
CFCF	0.4271	0.3453	0.3575	0.3596	0.3096	0.3534	0.3773
Staple	0.3965	0.2880	0.3538	0.3776	0.2275	0.3253	0.3327



A comparison of our approach (red) with ASRCF (yellow), HCFstar (blue), STRCF (green), and GFS-DCF (cyan) on ***Girl2*** and ***MotorRolling*** sequences from OTB100.