1 Department of Avionics, Indian Institute of Space Science and Technology, Thiruvananthapuram, Kerala, India - 695 547

2 Department of Electrical Engineering, Indian Institute of Technology, Tirupati, Andhra Pradesh, India - 517 506

Learning Rotation Adaptive Correlation Filters in Robust Visual Object Tracking

Asian Conference on Computer Vision

Litu Rout[1] Priya Mariam Raju[1] Deepak Mishra[1] **Rama Krishna Sai Subrahmanyam Gorthi**[2]

liturout1997@gmail.com, priyamariyam123@gmail.com, deepak.mishra@iist.ac.in, **rkg@iittp.ac.in**

July 1, 2019

Content

Introduction
- Contribution

Proposed Methodology
- Overall Framework
- Illumination Correction Filter
- Rotation Adaptive Correlation Filter
- False Positive Elimination
- Detection of Orientation
- Fast Sub-grid Detection
- Displacement Consistency

Experiments
- Ablation Studies
- Evaluation on VOT2016
- Evaluation on OTB100
- Evaluation of CF trackers
- Evaluation of Rotation Adaptive trackers

Concluding Remarks

Introduction

- Visual object tracking is one of the major challenges in the field of computer vision.
- Correlation Filter (CF) trackers are one of the most widely used categories in tracking.
- Tracking algorithms based on CFs still fail to efficiently detect the object in an unconstrained environment with dynamically changing object's representation.
- Visual object tracking with Correlation filters at a glance

- KCF framework demonstrates that the image translations can be modeled by circulant shifts.

Figure: Figure from [*]: Vertical cyclic shifts of a base sample. KCF Fourier domain formulation allows to train a tracker with all possible cyclic shifts of a base sample, both vertical and horizontal, without iterating them explicitly.

$$\varepsilon(f) = \|f(x_k) - y_k\|^2 + \lambda \|f\|^2 \quad (1)$$

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k^{xx}} + \lambda} \quad (2)$$

$$\hat{f} = \hat{\alpha .* k^{zx}} \quad (3)$$

[*] J. F. Henriques et. al., High speed tracking with kernelized correlation filters. TPAMI,2015.

Introduction

SRDCF*

- A spatial regularization component is introduced in the learning to penalize correlation filter coefficients depending on their spatial location.

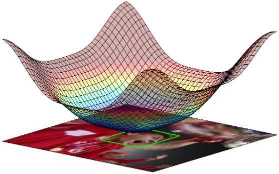

Figure: Visualization of the spatial regularization weights w employed in the learning of our SRDCF.

$$\varepsilon(f) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| w \cdot f^l \right\|^2 . \tag{4}$$

[*] Danelljan, M.et. al: Learning spatially regularized correlation filters for visual tracking. In: ICCV. (2015)

- An Illumination Correction filter (IC) is introduced in the tracking framework that eliminates the adverse effects of variable illuminations on feature extraction.
- We propose an approach to incorporate rotation adaptiveness in standard DCF by optimizing across the orientations of the target object in the detector stage. The orientation optimization helps in extracting robust features from properly oriented bounding boxes unlike most state-of-the-art trackers that rely on axis aligned bounding boxes.
- Building on it, we supervise the sub-grid localization cost function in the detector stage of DCF trackers. This cost function is intended to eliminate the false positives during detection.
- Further, we show the impact of enhancing smoothness through displacement correction, and demonstrate all these contributions on two popular CF trackers: Spatially Regularized Disriminative Correlation Filters, and Efficient Convolution Operators.

Proposed Methodology
Overall Framework

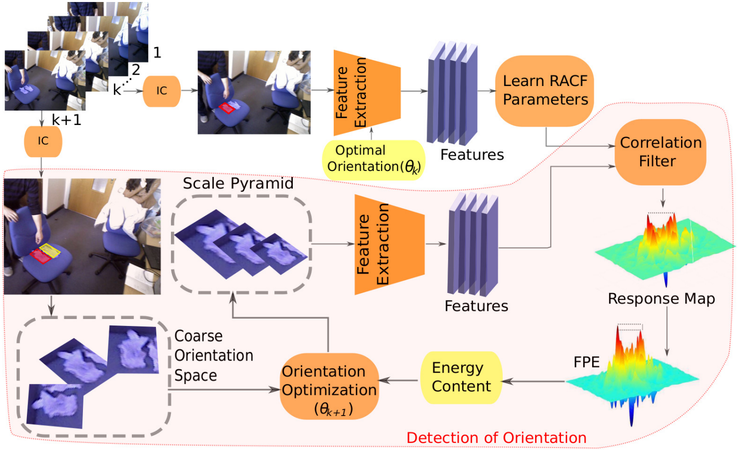

Figure: Overall pipeline of the proposed object tracking framework.

Illumination Correction

- At first, we employ a standard contrast stretching mechanism to adjust the intensities of each frame.
- The contrast stretched image is then subjected to unsharp masking, to suppress the low frequency interference, and enhance high variations.

This validates the fact that the robust feature extractors still lack high quality visual inputs, which otherwise can lead to substantial gain in performance.

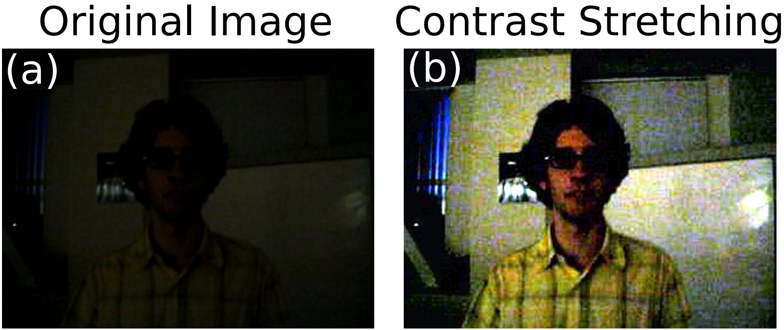

Figure: Qualitative analysis of Linear Contrast Stretching

Unsharp Masking

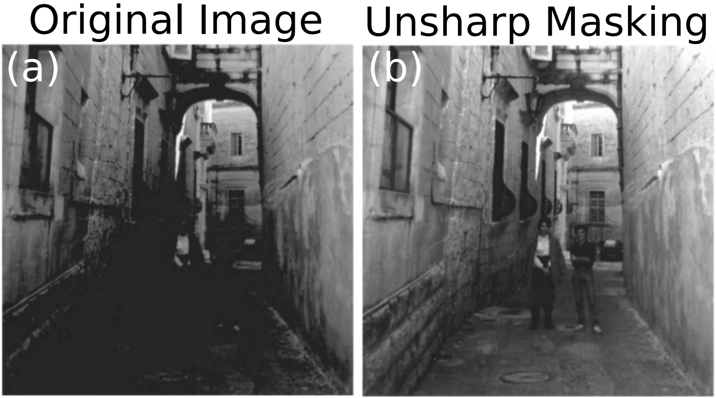

Figure: Qualitative analysis of Unsharp Masking

- Learning from properly oriented training samples $\left\{\left(x_k^\theta, y_k\right)\right\}_{k=1}^t$
- The convolution response $S_f(x_k^\theta)$ of the rotated training samples $x_k^\theta \in \mathbb{R}^d$ are computed by,

$$S_f(x_k^\theta) = \sum_{l=1}^{d} x_k^{\theta l} * f^l. \tag{5}$$

- The resulting cost function is given by,

$$\varepsilon_\theta\left(f\right) = \sum_{k=1}^{t} \alpha_k \left\| S_f(x_k^\theta) - y_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{w}{MN} \cdot f^l \right\|^2. \tag{6}$$

- The rotation adaptive filter parameters are computed by Gauss-Seidel iterative optimization,

$$\arg\min_{\hat{f}} \sum_{k=1}^{t} \alpha_k \left\| \sum_{l=1}^{d} \hat{x}_k^{\theta l} \cdot \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^{d} \left\| \frac{\hat{w}}{MN} * \hat{f}^l \right\|^2. \tag{7}$$

- The convolution response of the test sample is computed by,

$$\hat{s}_\theta(u, v) := \mathcal{F}\left\{S_f(z^\theta)\right\} = \sum_{l=1}^{d} \hat{z}^{\theta l} \cdot \hat{f}^l. \tag{8}$$

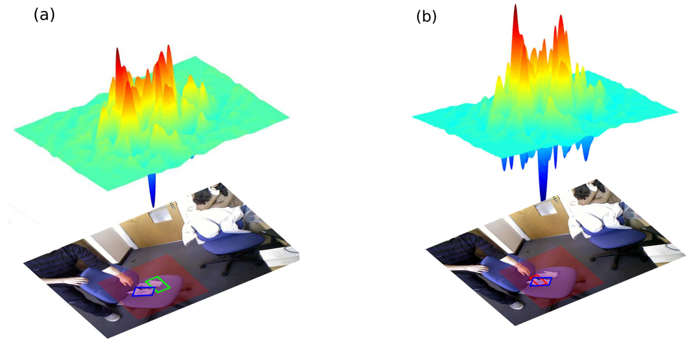

Figure: Convolution Response Map. Blue, Green, and Red bounding boxes represent ground truth, SRDCF, and RIDF-SRDCF(our) output, respectively.

- Maximize probability of detection subject to minimum deviation from previous centroid, $\frac{s(u,v)}{\left\|\left(u-u_k^*, v-v_k^*\right)\right\|}$.

Proposed Methodology
Detection of Orientation

- The aim is to find orientation that maximizes the total energy content in the convolution response map by,

$$\theta_{k+1} = \arg\max_{\theta \in \Phi} \left\{ \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \left(\frac{S_\theta(u,v)}{\left\| \left(u - u_k^*, v - v_k^*\right) \right\|} \right)^2 \right\}. \quad (9)$$

Here, $\Phi := \{\theta_k \pm a\delta\}$, where $a = 0, 1, 2, \ldots, A$.

- The computational complexity of standard SRDCF (7 FPS) is given by,

$$\mathcal{O}\left(SdMN \log MN + SMNN_{Ne} + \left(d + k^2\right) dMNN_{GS}\right). \quad (10)$$

- The computational complexity of our RIDF-SRDCF (5 FPS) is given by,

$$\mathcal{O}\left(ASdMN \log MN + \left(ASdMN + ASMN\right) N_{Ne} + MN + \left(d + k^2\right) dMNN_{GS}\right). \quad (11)$$

- Incorporate the FPE and DoO techniques in Fast Sub-grid Detection method of standard SRDCF formulation.
- Apply the Newton's optimization strategy, as in SRDCF, for finding the sub-grid location that maximizes the detection score.
- Thus, we compute the sub-grid location that corresponds to maximum detection score by,

$$\left(u_{k+1}^{*}, v_{k+1}^{*}\right) = \arg\max_{(u,v)\in[0,M)\times[0,N)} \left\{ \frac{S_{\theta_{k+1}}(u,v)}{\left\|\left(u-u_k^{*}, v-v_k^{*}\right)\right\|} \right\}, \tag{12}$$

starting at $(u^{(0)}, v^{(0)}) \in \Omega$, such that $\left\{ \frac{S_{\theta_{k+1}}(u^{(0)},v^{(0)})}{\left\|\left(u^{(0)}-u_k^{*}, v^{(0)}-v_k^{*}\right)\right\|} \right\}$ is maximal.

- Motivated by the displacement consistency techniques*, we enhance the degree of smoothness imposed on the movement variables, such as speed and angular displacement. We update the sub-grid location, $\left(u_{k+1}^{*}, v_{k+1}^{*}\right)$ obtained from equation (12) by,

$$\begin{aligned} \left(u_{k+1}^{*}, v_{k+1}^{*}\right) &= \left(u_{k}^{*}, v_{k}^{*}\right) + d_{1n}\angle\varphi_{1n}, \\ d_{1n} &= \omega_d \times d_1 + (1-\omega_d) \times d_0, \\ \varphi_{1n} &= \omega_a \times \varphi_1 + (1-\omega_a) \times \varphi_0, \end{aligned} \tag{13}$$

where,
$d_0 = \left\|\left(u_k^* - u_{k-1}^*, v_k^* - v_{k-1}^*\right)\right\|$, $d_1 = \left\|\left(u_{k+1}^* - u_k^*, v_{k+1}^* - v_k^*\right)\right\|$,
$\varphi_0 = \arctan\left(u_k^* - u_{k-1}^*, v_k^* - v_{k-1}^*\right)$,
$\varphi_1 = \arctan\left(u_{k+1}^* - u_k^*, v_{k+1}^* - v_k^*\right), \omega_d = 0.9, \omega_a = 0.9.$

- This enforce motion consttency in physical units rather than on position vector.

[*] Litu R et. al., “ Rotation Adaptive Visual Object Tracking with Motion Consistency”, WACV, 2018.

- We progressively integrate Displacement consistency (D), False positive elimination (F), Rotation adaptiveness (R), Illumination correction (I), and their combinations into ECO framework for faster experimentaion, and assimilate the impact of each individual component on AEO, which is the standard metric on VOT2016 benchmark.
- FPE scheme improves the performance in every integration, and illumination correction alone provides a gain of 7.7% over base RDF-ECO

Table: Quantitative evaluation of Ablative trackers on a set of 16 challenging videos from VOT2016 benchmark.

Tracker	ECO	D-ECO	DF-ECO	R-ECO	RF-ECO	RD-ECO	RDF-ECO	RIDF-ECO
AEO	0.357	0.360	0.362	0.383	0.386	0.395	0.402	0.433
%Gain	Baseline	0.8	1.4	7.3	8.1	10.6	12.6	21.3

- As per the results in Table 2, the I-SRDCF, RDF-SRDCF, and RIDF-SRDCF provide a considerable improvement of 3.53%, 10.60%, and 11.41% in AEO, 4.83%, 17.87%, and 13.04% in robustness, respectively.
- The RIDF-ECO performs favourably against the state-of-the-art trackers including MDNet (won VOT2015) and CCOT (won VOT2016) with a slight improvement of 1.71% in AEO, and as high as 6.41% in robustness.

Table: State-of-the-art comparison on whole VOT2016 dataset.

Trackers	SRDCF	I-SRDCF	RDF-SRDCF	**RIDF-SRDCF**	TCNN	CCOT	ECO	MDNet	**RIDF-ECO**
AEO	0.1981	0.2051	0.2191	0.2207	0.3249	0.3310	0.3563	0.3584	0.3624
Failure Rate (Robustness)	2.07	1.97	1.70	1.80	0.96	0.83	0.78	0.76	0.73

1

[1] The percentage gain is computed relative to baseline.

Table: State-of-the-art comparison on OTB100 dataset.

Trackers	**RIDF-ECO**	ECO	MDNet	CCOT	**RIDF-SRDCF**	DeepSRDCF	SRDCF	CFNet	Staple	KCF
Out-of-view	0.767	0.726	0.708	0.725	0.712	0.619	0.555	0.423	0.518	0.550
Occlusion	0.721	0.710	0.702	0.692	0.652	0.625	0.641	0.573	0.610	0.535
Illumination Variation	0.702	0.662	0.688	0.676	0.649	0.631	0.620	0.561	0.601	0.530
Low Resolution	0.734	0.652	0.663	0.642	0.588	0.438	0.537	0.545	0.494	0.384
Background Clutter	0.648	0.638	0.697	0.620	0.634	0.616	0.612	0.592	0.580	0.557
Deformation	0.687	0.687	0.722	0.657	0.652	0.645	0.641	0.618	0.690	0.608
In-plane rotation	0.696	0.645	0.656	0.653	0.635	0.625	0.615	0.606	0.596	0.510
Out-of-plane rotation	0.682	0.665	0.707	0.663	0.646	0.637	0.618	0.593	0.594	0.514
Fast Motion	0.716	0.698	0.671	0.694	0.693	0.640	0.599	0.547	0.526	0.482
Overall Success Rate	0.702	0.691	0.678	0.671	0.641	0.635	0.598	0.589	0.581	0.477
Overall Precision	0.937	0.910	0.909	0.898	0.870	0.851	0.789	0.777	0.784	0.696

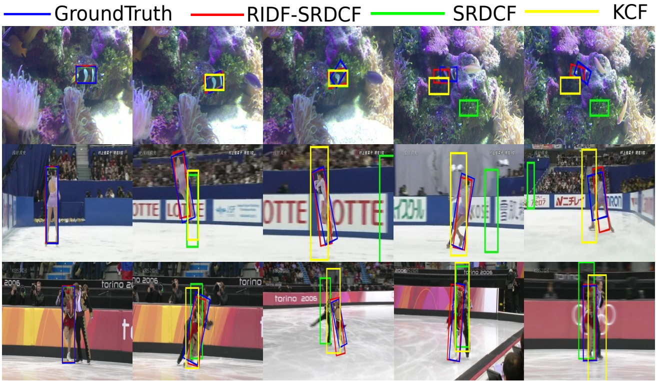

Figure: Qualitative analysis of RIDF-SRDCF. The proposed tracker successfully tracks the target under severe rotation, unlike SRDCF and KCF, estimate the orientation of the target object, thus leads to substantial gain in overall performance.

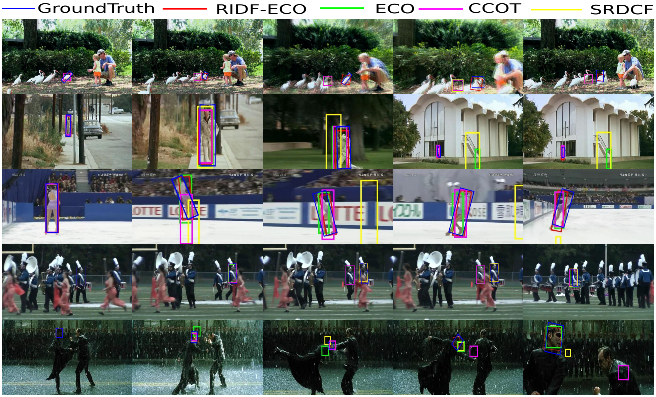

Figure: Qualitative analysis of RIDF-SRDCF. The proposed tracker successfully tracks the target under severe rotation.

Video demonstration of Qualitative Analysis

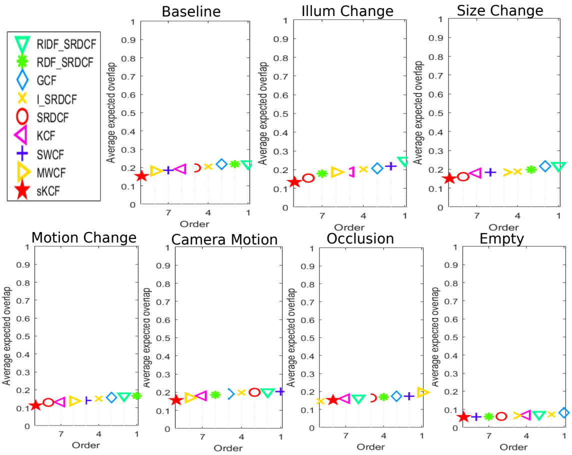

Figure: Average Expected Overlap analysis of correlation filter based trackers.

Experiments

Evaluation of Rotation Adaptive trackers

- We compare our rotation adaptive scheme with two recent approaches that aims at addressing this issue heuristically.
- As per our experiments, we report that the proposed rotation adaptive scheme outperforms these counterparts on VOT2016 benchmark. Since base CF trackers are used as core components in most trackers, we believe that the proposed performance gain will be reflected positively in all their derivatives.

Table: Comparison with two recent rotation adaptive trackers on VOT2016.

Trackers	RAJSSC	SRDCF[1]	SiameseFC-DSR	**RIDF-SRDCF**	ECO[2]	**RIDF-ECO**
AEO	0.1664	0.1981	0.2084	0.2207	0.3563	0.3624

[1] Danelljan, M., Häger, G., Shahbaz Khan, F., Felsberg, M.: Learning spatially regularized correlation filters for visual tracking. In: ICCV. (2015)
[2] Danelljan, M., Bhat, G., Shahbaz Khan, F., Felsberg, M.: ECO: efficient convolu- tion operators for tracking. In: CVPR. (2017)

Concluding Remarks

- We demonstrated effective image enhancement prior to (deep) feature extraction, can yield considerable gain in tracking paradigm.

Concluding Remarks

- We demonstrated effective image enhancement prior to (deep) feature extraction, can yield considerable gain in tracking paradigm.
- We proposed rotation adaptive search in correlation filter with in standard DCF formulation, that greatly improve the performance.

Concluding Remarks

- We demonstrated effective image enhancement prior to (deep) feature extraction, can yield considerable gain in tracking paradigm.
- We proposed rotation adaptive search in correlation filter with in standard DCF formulation, that greatly improve the performance.
- We renovated the sub-grid detection approach by optimizing object's orientation with velocity smoothness through false positive elimination.

Concluding Remarks

- We demonstrated effective image enhancement prior to (deep) feature extraction, can yield considerable gain in tracking paradigm.
- We proposed rotation adaptive search in correlation filter with in standard DCF formulation, that greatly improve the performance.
- We renovated the sub-grid detection approach by optimizing object's orientation with velocity smoothness through false positive elimination.
- Also, the supervision of displacement consistency on CF trackers showed promising results in various scenarios.

Concluding Remarks

- We demonstrated effective image enhancement prior to (deep) feature extraction, can yield considerable gain in tracking paradigm.
- We proposed rotation adaptive search in correlation filter with in standard DCF formulation, that greatly improve the performance.
- We renovated the sub-grid detection approach by optimizing object's orientation with velocity smoothness through false positive elimination.
- Also, the supervision of displacement consistency on CF trackers showed promising results in various scenarios.
- **Moreover, since the DCF formulation is used as backbone of most state-of-the-art trackers, we believe that the proposed rotation adaptive scheme in correlation filters can be suitably integrated into many frameworks and will be useful in boosting the tracking research forward.**

Thank You for Your Attention!!