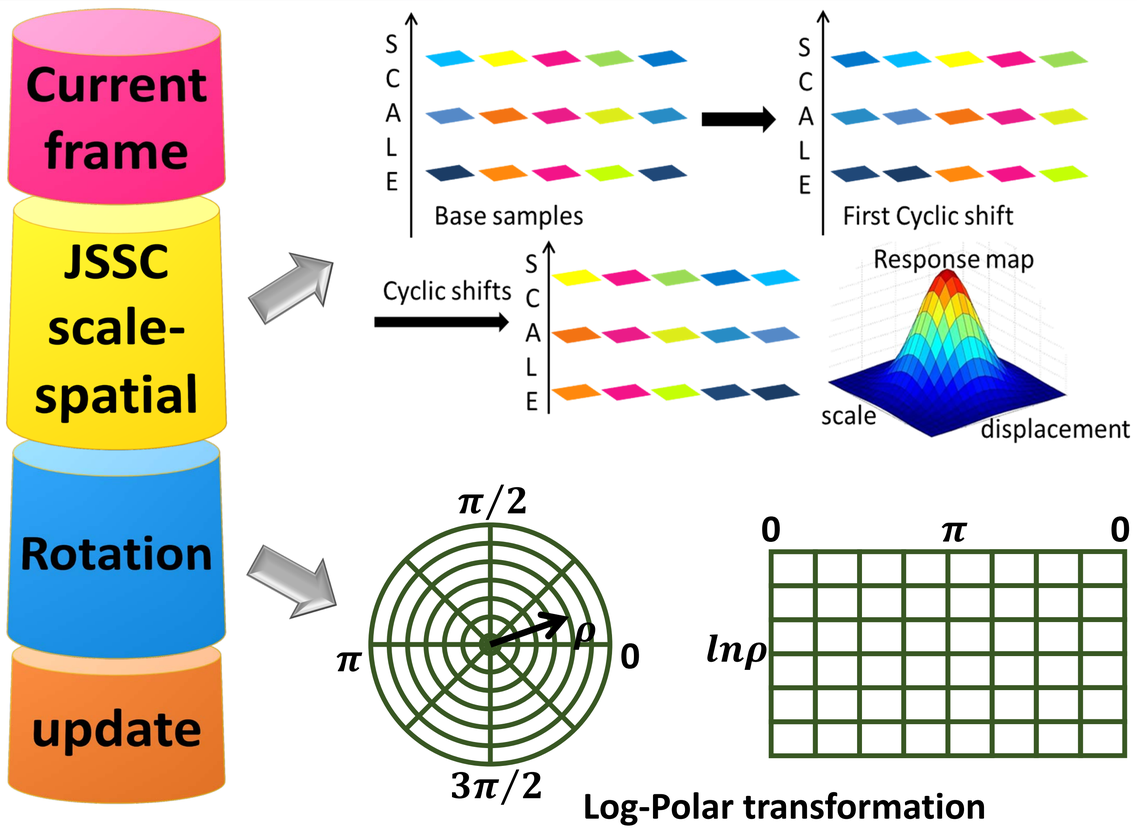
Joint Scale-Spatial Correlation Tracking with Adaptive Rotation Estimation

Mengdan Zhang, Junliang Xing, Jin Gao, Xinchu Shi, Qiang Wang, Weiming Hu
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
No. 95, Zhongguancun East Road, Beijing 100190, P. R. China

Abstract

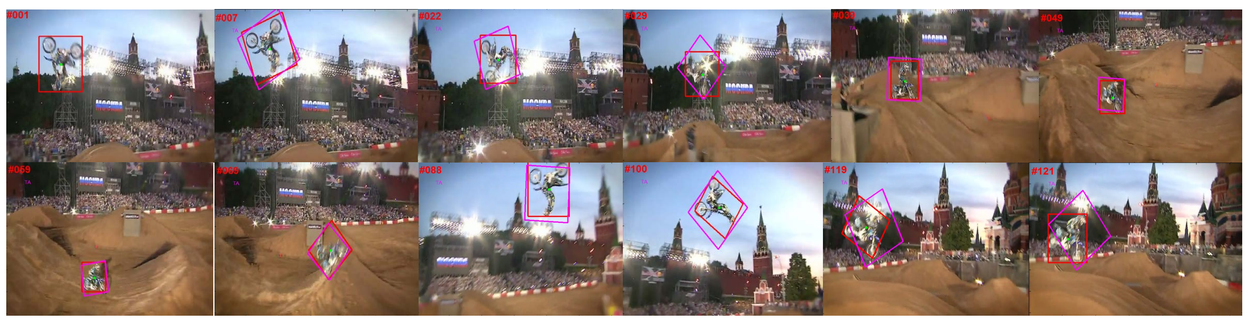
Boosted by large and standardized benchmark datasets, visual object tracking has made great progress in recent years and brought about many new trackers. Among these trackers, correlation filter based tracking schema exhibits impressive robustness and accuracy. In this work, we present a fully functional correlation filter based tracking algorithm which is able to simultaneously model target appearance changes from spatial displacements, scale variations, and rotation transformations. The proposed tracker first represents the exhaustive template search in the joint scale and spatial space by a block-circulant matrix. Then, by transferring the target template from the Cartesian coordinate system to the Log-Polar coordinate system, the circulant structure is well preserved for the target even after whole orientation rotation. With these novel representation and transformation, object tracking is efficiently and effectively performed in the joint space with fast Fourier Transform. Experimental results on the VOT2015 benchmark dataset demonstrate its superior performance over state-of-the-art tracking algorithms.

Block Diagram of the Proposed Tracker RAJSSC

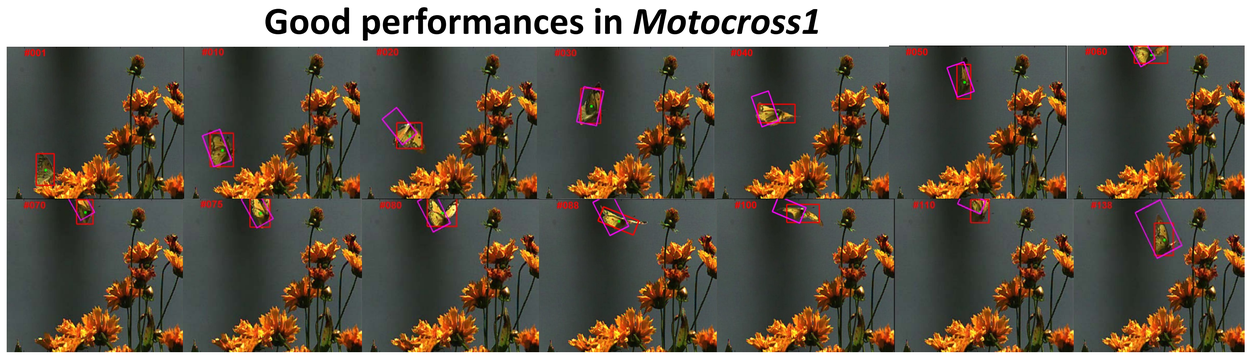


Log-Polar transformation

Scale and rotation estimation



Good performances in *Motocross1*



Unsatisfactory performances in *Butterfly*

The red and magenta bounding boxes denote the ground-truths and our tracking results respectively.

JSSC (Joint scale-spatial correlation filter)

For simplicity, assume a 1D image and a single channel feature. The JSSC tracker is trained using S base samples of size $1 \times N$ obtained from the recent scale level and neighboring levels. Taking advantages of the cyclic property and appropriate padding, it considers all cyclic shifts $\{x_s(n)\}, s \in \{1,2,\cdots,S\}, n \in \{0,1,\cdots,N-1\}$ as the training samples for the target template estimation. The matching scores $\boldsymbol{y}$ obey a multivariate Gaussian distribution in the joint scale-spatial space. Then, we have:

$$\min_{\omega} \sum_{n,s} |\langle \phi(x_s(n)), \omega \rangle - y_s(n)|^2 + \lambda \|\omega\|^2$$

The JSSC solution in the Fourier domain is extended as

$$\hat{\vec{\alpha}}^* = \left(diag\left(g(u_0), g(u_1), \cdots, g(u_{N-1})\right) + \lambda I_{SN}\right)^{-1} \hat{\vec{\boldsymbol{y}}}^*$$

$$g(u_c) = \begin{bmatrix} \hat{k}_{x1x1,c} & \cdots & \hat{k}_{x1xS,c} \\ \vdots & \ddots & \vdots \\ \hat{k}_{xSx1,c} & \cdots & \hat{k}_{xSxS,c} \end{bmatrix}$$

where $k_{xixj,c}$ is the c-th element of the base sample of the Gaussian kernel matrix K_{ij}, the horizontal bars represent the rearrangement to obtain the block-circulant matrices, the hats represent the Fourier Transform.

In the tracking section, the candidates Z to be matched with the target template are extracted in the same way from the joint scale-spatial space. The matching scores can be evaluated via

$$f(Z) = K^{ZX}\alpha$$

Considering the block-circulant matrix properties, the full tracking response is given by

$$\hat{\vec{f}}(Z) = diag(h^*(u_0), h^*(u_1), \cdots, h^*(u_{N-1}))\hat{\vec{\alpha}}$$

$$h(u_c) = \begin{bmatrix} \hat{k}_{z1x1,c} & \cdots & \hat{k}_{z1xS,c} \\ \vdots & \ddots & \vdots \\ \hat{k}_{zSx1,c} & \cdots & \hat{k}_{zSxS,c} \end{bmatrix}$$

Moreover, linear interpolation is adopted according to the tracking response to ensure the continuity of the scale and position estimation.

Adaptive rotation estimation

We propose to perform rotation estimation using a unified correlation tracking framework by taking the Log-Polar transformation. The rotation template can be trained on all the cyclic shift versions of x_r, denoted by $x_r(\theta), \theta \in \{1,2,\cdots,R\}$. The sample interval is $\Delta = \frac{2\pi}{R}$. Each sample is also assigned with a score generated by a Gaussian function $\boldsymbol{y_r}$. Similarly, by minimizing the regression error, we get the solution via

$$\hat{\alpha}_r = \frac{\hat{y}_r}{\hat{k}_{x_r x_r} + \lambda}$$

where $\hat{k}_{x_r x_r}$ is the Fourier Transform of the vector whose i-th element is $\kappa(x_r(i), x_r)$. In visual tracking scenarios, the template matching scores are calculated as

$$y_r = \mathcal{F}^{-1}\left(\hat{k}_{x_r z_r} \odot \hat{\alpha}_r\right)$$

Experiment 1: The Effectiveness of Our Approach

To evaluate the performance gain of our fully functional correlation filter based tracker, we compare it with four variants of correlation filter based trackers on the VOT2014 benchmark including KCF [12], KCF14, an enhanced KCF with scale estimation [15], SAMF [17], and DSST [5].

Table 1. Comparisons of accuracy and robustness among correlation filter-based trackers in the baseline experiment.

	accuracy (overlap ratio)					robustness (failure times)				
	KCF	KCF14	SAMF	DSST	RAJSSC	KCF	KCF14	SAMF	DSST	RAJSSC
ball	0.702	0.758	0.775	0.568	0.767	1	1	1	1	0
basketball	0.574	0.645	0.751	0.638	0.621	2	0	0	1	0
bicycle	0.454	0.630	0.618	0.583	0.712	1	0	0	0	0
bolt	0.522	0.490	0.562	0.562	0.705	3	3	2	1	0
car	0.421	0.713	0.512	0.742	0.734	0	0	0	0	0
david	0.746	0.822	0.818	0.807	0.796	0	0	0	0	0
diving	0.233	0.255	0.246	0.442	0.288	5	4	4	1	5
drunk	0.434	0.536	0.569	0.551	0.576	0	0	0	0	0
fernando	0.402	0.411	0.395	0.340	0.468	1	1	1	1	1
fish1	0.438	0.419	0.496	0.321	0.436	3	3	3	1	5
fish2	0.299	0.266	0.299	0.353	0.432	4	6	5	4	2
gymnastics	0.528	0.537	0.538	0.632	0.582	3	1	2	5	1
hand1	0.389	0.563	0.547	0.215	0.596	6	3	3	2	2
hand2	0.438	0.498	0.465	0.528	0.550	8	6	5	6	2
jogging	0.760	0.799	0.822	0.790	0.534	1	1	1	1	1
motocross	0.372	0.366	0.402	0.421	0.661	5	2	4	4	1
polarbear	0.662	0.780	0.709	0.635	0.712	0	0	0	0	0
skating	0.488	0.677	0.452	0.586	0.645	0	1	0	0	0
sphere	0.713	0.90	0.880	0.927	0.738	0	0	0	0	0
sunshade	0.761	0.763	0.759	0.783	0.773	0	0	0	0	0
surfing	0.797	0.805	0.804	0.906	0.821	0	0	0	0	0
torus	0.757	0.857	0.841	0.811	0.791	0	0	0	0	0
trellis	0.546	0.798	0.825	0.808	0.817	0	0	0	0	0
tunnel	0.318	0.687	0.553	0.812	0.718	0	0	0	0	0
woman	0.755	0.744	0.761	0.790	0.653	2	1	1	1	1
MEAN	0.540	0.629	0.616	0.622	0.645	1.80	1.32	1.28	1.16	0.84

The table below shows the exact per-visual attribute normalized accuracy and robustness ranks of the top twenty trackers.

. VOT2014 competition report. The top, second and third lowest average ranks are shown in red, blue and green respectively.

	baseline		region_noise				
	Acc. Rank	Rob. Rank	Acc. Rank	Rob. Rank	Acc. Rank	Rob. Rank	Rank
RAJSSC	4.96	11.49	6.17	11.41	5.56	11.45	8.51
DSST	5.99	12.52	5.96	12.87	5.98	12.69	9.33
SAMF	5.76	14.34	5.65	12.78	5.70	13.56	9.63
KCF	5.51	15.41	5.58	13.05	5.54	14.23	9.89
PLT_14	14.72	6.44	13.74	5.01	14.23	5.73	9.98
DGT	11.67	9.55	9.15	10.11	10.41	9.83	10.12
PLT_13	18.34	3.83	17.38	4.83	17.86	4.33	11.10
eASMS	14.22	13.87	11.42	14.34	12.82	14.11	13.46
HMMTxD	10.28	20.77	9.81	19.57	10.05	20.17	15.11
MCT	16.88	14.08	17.58	13.00	17.23	13.54	15.38
ACAT	13.84	15.23	17.81	14.96	15.82	15.10	15.46
MatFlow	22.00	8.88	19.14	14.64	20.57	11.76	16.17
ABS	20.68	18.62	15.44	15.29	18.06	16.95	17.51
ACT	20.85	16.63	22.27	15.22	21.56	15.92	18.74
qwsEDFT	17.58	19.45	18.64	21.06	18.11	20.25	19.18
LGTv1	29.23	11.78	26.45	9.39	27.84	10.59	19.21
VTDMG	21.48	18.40	20.65	16.98	21.06	17.69	19.38
BDF	23.17	17.93	21.74	18.15	22.45	18.04	20.25
Struck	20.87	21.12	21.51	18.85	21.19	19.99	20.59
DynMS	22.82	19.47	21.51	19.51	22.16	19.49	20.83

Experiment 2: VOT2015 challenge results

Name	Acc. Rank	Overlap ratio	Rob. Rank	Failures	Name	Acc. Rank	Overlap ratio	Rob. Rank	Failures
RAJSSC	1.15	0.52	1.11	1.63	NCC	1.88	0.36	1.89	10.74

National Laboratory of Pattern Recognition
Institute of Automation Chinese Academy of Sciences, P.R.C
Contact information: mengdan.zhang@nlpr.ia.ac.cn