Supplementary Document For
IGSSTRCF: Importance Guided Sparse Spatio-Temporal Regularized Correlation Filters For Tracking

Monika Jain
QUT, Brisbane, Australia
IIIT, Delhi, India
monikaj@iiitd.ac.in

A V Subramanyam
IIIT, Delhi, India
subramanyam@iiitd.ac.in

Simon Denman
QUT, Brisbane, Australia
s.denman@qut.edu.au

Sridha Sridharan
QUT, Brisbane, Australia
s.sridharan@qut.edu.au

Clinton Fookes
QUT, Brisbane, Australia
c.fookes@qut.edu.au

1. VOT Toolkit Results for VOT-2019 Dataset

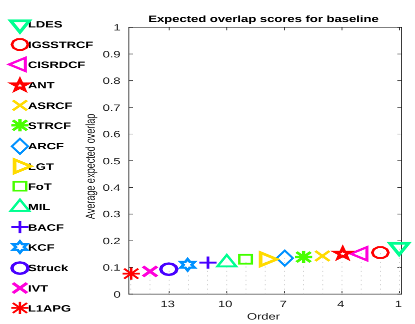

Figure 1: Expected overlap scores for the baseline experiments on VOT-2019 [8], showing the that proposed IGSSTRCF tracker performs the second best

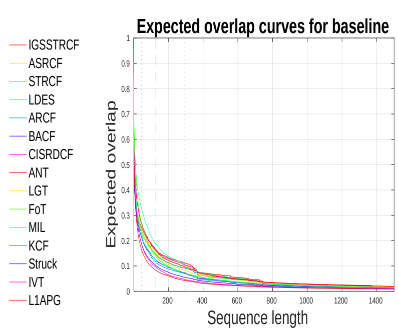

Figure 2: Expected overlap curves for the baseline experiments on VOT-2019 [8]

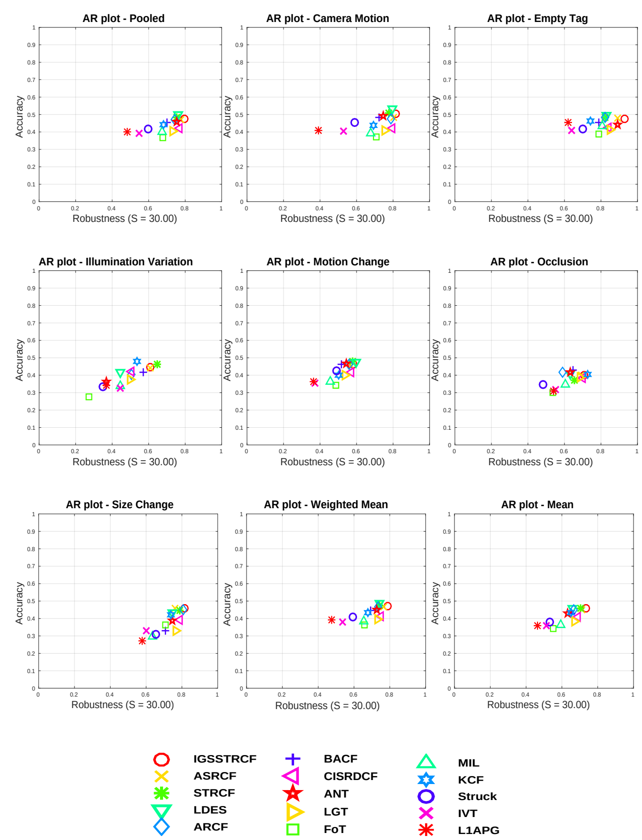

Figure 3: AR plots of individual challenges for the baseline experiments on VOT-2019 [8]. The proposed IGSSTRCF tracker is compared with recent state-of-the-art trackers

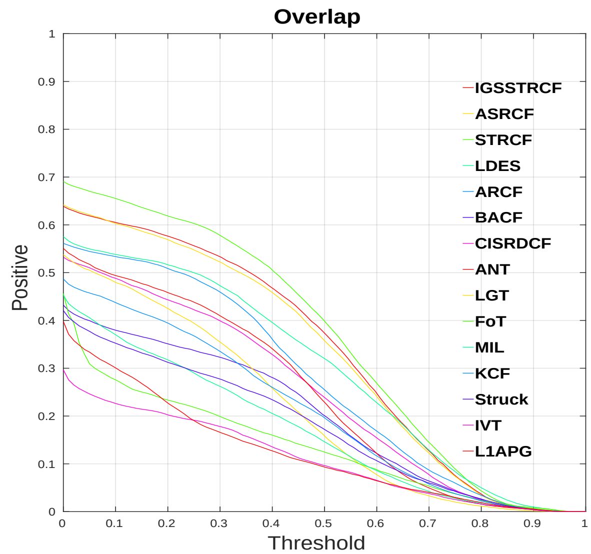

Figure 4: Expected Overlap curves for unsupervised experiments on VOT-2019 [8]

Table 1: Accuracy comparison of the proposed IGSSTRCF approach with recent trackers for the baseline experiments on VOT-2019 [8]. Here CM = Camera Motion, EMP = Empty Tag, IV= Illumination Variation, MC = Motion Change, OCC = Occlusion, SC = Size Change, and Wt. Mean = Weighted Mean. The top three trackers are shown in **red**, **blue** and **green**

	CM	**EMP**	**IV**	**MC**	**OCC**	**SC**	**Mean**	**Wt. Mean**	**Pooled**
IGSSTRCF	**0.5037**	0.4735	**0.4451**	0.4625	0.4014	**0.4596**	**0.4577**	**0.4730**	**0.4760**
ASRCF [1]	0.4825	**0.4777**	**0.4432**	0.4524	0.4003	**0.4565**	**0.4521**	0.4652	0.4707
STRCF [9]	0.4780	0.4540	0.4351	0.4444	0.3741	0.4451	0.4381	0.4523	0.4571
LDES [10]	**0.5321**	**0.4972**	0.4177	**0.4745**	0.4071	0.4330	**0.4603**	**0.4882**	**0.4986**
ARCF [5]	0.4762	**0.4842**	0.4050	**0.4637**	**0.4151**	**0.4515**	0.4493	**0.4669**	**0.4716**
BACF [6]	0.4821	0.4546	0.4183	0.4607	**0.4306**	0.3290	0.4292	0.4476	0.4533
CISRDCF [8]	0.4201	0.4243	0.4224	0.4181	0.3853	0.3919	0.4104	0.4147	0.4198
ANT [8]	**0.4906**	0.4419	0.3650	**0.4648**	**0.4187**	0.3881	0.4282	0.4518	0.4581
LGT [8]	0.4067	0.4121	0.3752	0.3991	0.3879	0.3282	0.3849	0.3960	0.4030
FoT [8]	0.3709	0.3867	0.2773	0.3409	0.3020	0.3650	0.3405	0.3621	0.3657
MIL [8]	0.3940	0.4329	0.3391	0.3651	0.3445	0.2983	0.3623	0.3847	0.3985
KCF [8]	0.4363	0.4627	**0.4809**	0.3996	0.4056	0.4214	0.4344	0.4348	0.4409
Struck [8]	0.4530	0.4191	0.3351	0.4253	0.3452	0.3082	0.3810	0.4103	0.4172
IVT [8]	0.4055	0.4101	0.3245	0.3566	0.3161	0.3310	0.3573	0.3811	0.3908
L1APG [8]	0.4095	0.4562	0.3409	0.3634	0.3087	0.2716	0.3584	0.3901	0.4001

Table 2: Robustness comparison of the proposed IGSSTRCF approach with recent trackers for the baseline experiments on VOT-2019 [8], following the same abbreviations as Table 1. The top three trackers are shown in **red**, **blue** and **green**

	CM	EMP	**IV**	**MC**	**OCC**	**SC**	**Mean**	**Wt. Mean**	**Pooled**
IGSSTRCF	**56.00**	**18.00**	**8.00**	**69.00**	**26.00**	**19.00**	**32.66**	**39.00**	**152.00**
ASRCF [1]	**60.00**	**28.00**	**8.00**	**72.00**	28.00	25.00	**36.83**	**44.58**	**167.00**
STRCF [9]	96.00	61.00	12.00	87.00	41.00	28.00	54.16	70.02	243.00
LDES [10]	**62.00**	45.00	13.00	**66.00**	35.00	27.00	41.33	50.27	182.00
ARCF [5]	65.00	47.00	11.00	74.00	40.00	**20.00**	42.83	52.71	197.00
BACF [6]	88.00	58.00	**9.00**	84.00	33.00	32.00	50.66	65.70	238.00
CISRDCF [8]	**62.00**	42.00	11.00	**72.00**	**27.00**	**22.00**	**39.33**	**48.98**	**176.00**
ANT [8]	80.00	**28.00**	16.00	78.00	35.00	27.00	44.00	53.09	187.00
LGT [8]	77.11	**39.44**	11.20	79.80	28.65	**24.80**	43.50	54.86	206.85
FoT [8]	94.00	58.00	21.00	92.00	47.00	32.00	57.33	70.43	258.00
MIL [8]	105.00	52.00	13.00	100.00	38.00	42.00	58.33	73.65	261.00
KCF [8]	99.00	73.00	**10.00**	88.00	**24.00**	28.00	53.66	73.09	255.00
Struck [8]	143.00	87.00	17.00	91.00	55.00	39.00	72.00	96.32	344.00
IVT [8]	172.00	109.00	13.00	127.00	45.00	47.00	85.50	117.77	399.00
L1APG [8]	253.73	116.20	16.00	129.53	46.93	50.86	102.21	147.77	482.40

Table 3: Overlap AUC comparison of the proposed IGSSTRCF approach with recent trackers for unsupervised experiments on VOT-2019 [8], following the same abbreviations as Table 1. The top three trackers are shown in **red**, **blue** and **green**

	CM	**EMP**	**IV**	**MC**	**OCC**	**SC**	**OverAll**
IGSSTRCF	**0.3542**	**0.3286**	**0.2453**	**0.3178**	**0.2475**	**0.3357**	**0.3286**
ASRCF [1]	**0.3446**	**0.3268**	**0.2226**	**0.3113**	**0.2416**	**0.3303**	**0.3230**
STRCF [9]	0.3001	**0.3341**	0.2161	0.2481	0.1879	**0.3162**	**0.2980**
LDES [10]	**0.3131**	0.3231	**0.2693**	0.2457	0.1887	0.2305	0.2940
ARCF [5]	0.2923	0.2717	0.1873	**0.2536**	0.1949	0.2844	0.2690
BACF [6]	0.2171	0.2049	0.1779	0.1797	0.1787	0.1297	0.1959
CISRDCF [8]	0.2573	0.2382	0.1920	0.2054	**0.2046**	0.2055	0.2417
ANT [8]	0.2934	0.2088	0.1674	0.2269	0.1871	0.2210	0.2390
LGT [8]	0.2595	0.1981	0.1410	0.1773	0.1254	0.1968	0.2062
FoT [8]	0.1245	0.1668	0.0586	0.0886	0.0799	0.1365	0.1354
MIL [8]	0.1694	0.1931	0.1450	0.1165	0.1291	0.1279	0.1664
KCF [8]	0.1981	0.2193	0.2008	0.1847	0.1855	0.2131	0.2059
Struck [8]	0.1857	0.1910	0.1351	0.1638	0.1192	0.1304	0.1743
IVT [8]	0.1050	0.1288	0.0442	0.0663	0.0631	0.1333	0.1095
L1APG [8]	0.0793	0.1844	0.1276	0.0783	0.0953	0.0883	0.1224

2. VOT Toolkit Results for VOT-2017 Dataset

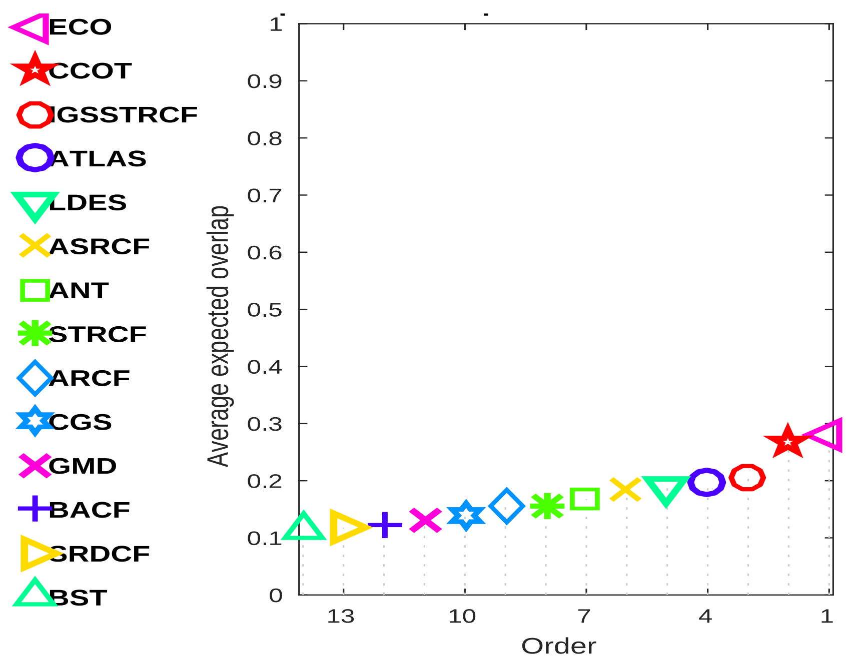

Figure 5: Expected overlap scores for the baseline experiments on VOT-2017 [7], showing the that proposed IGSSTRCF tracker outperforms several state-of-the-art trackers

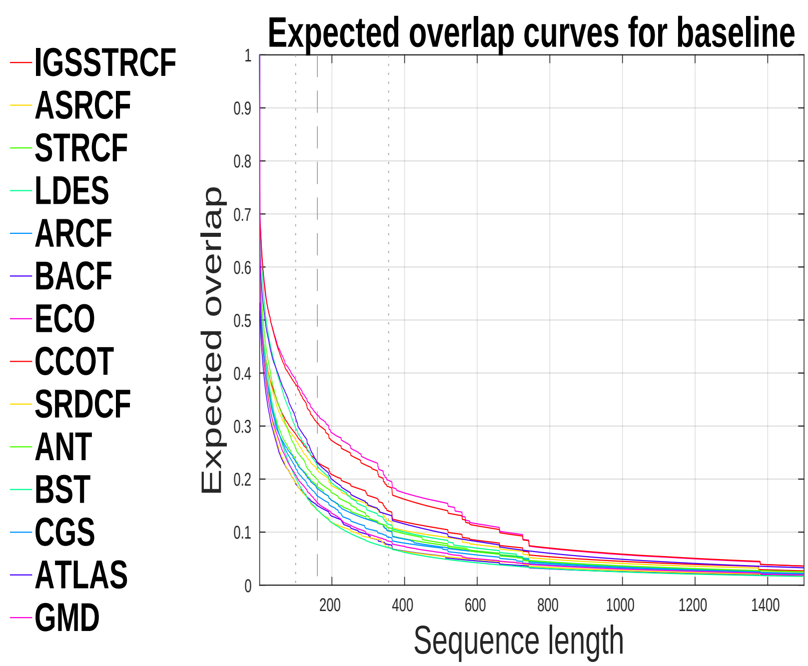

Figure 6: Expected overlap curves for the baseline experiments on VOT-2017 [7]

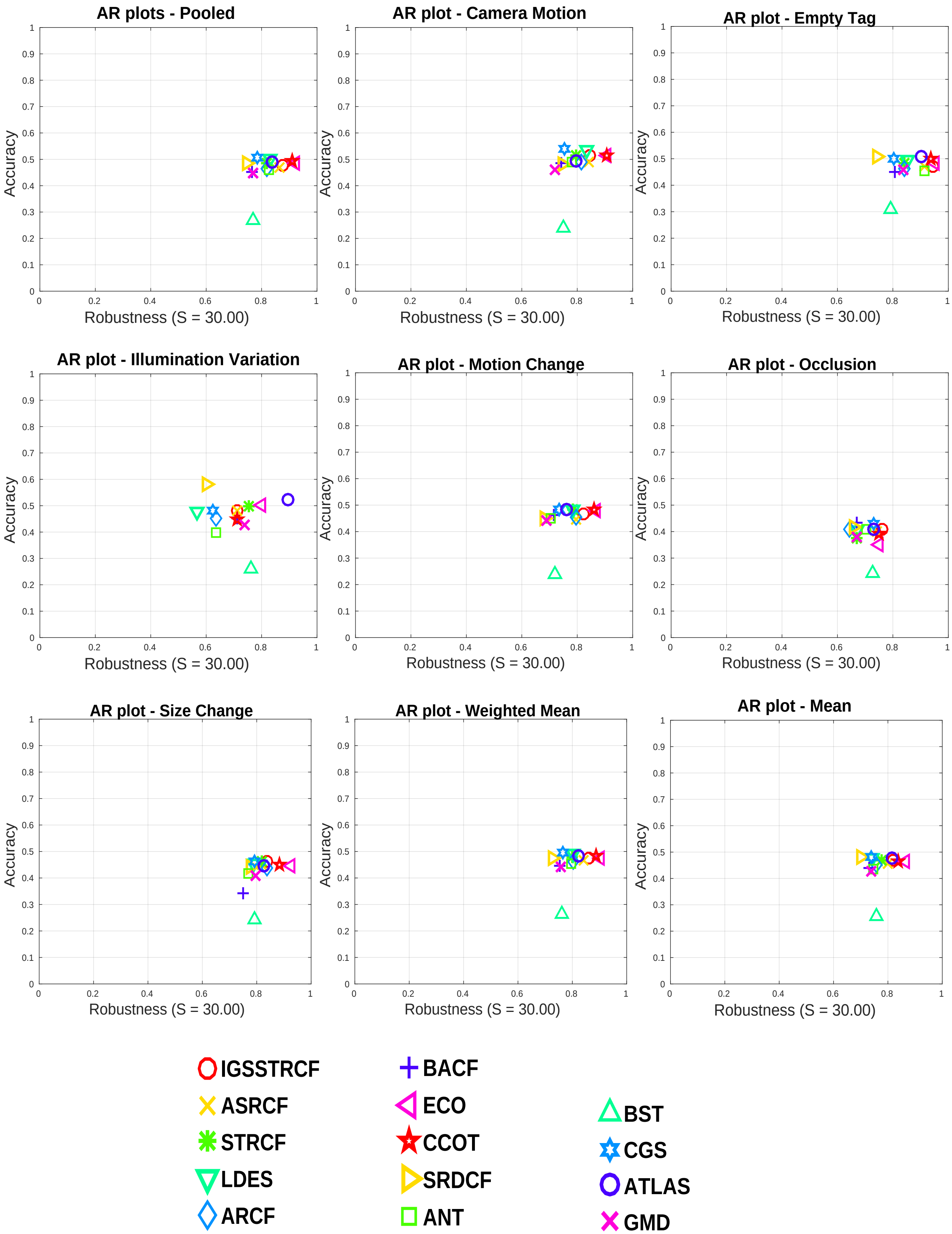

Figure 7: AR plots of individual challenges for the baseline experiments on VOT-2017 [7]. The proposed IGSSTRCF tracker is compared with recent state-of-the-art trackers

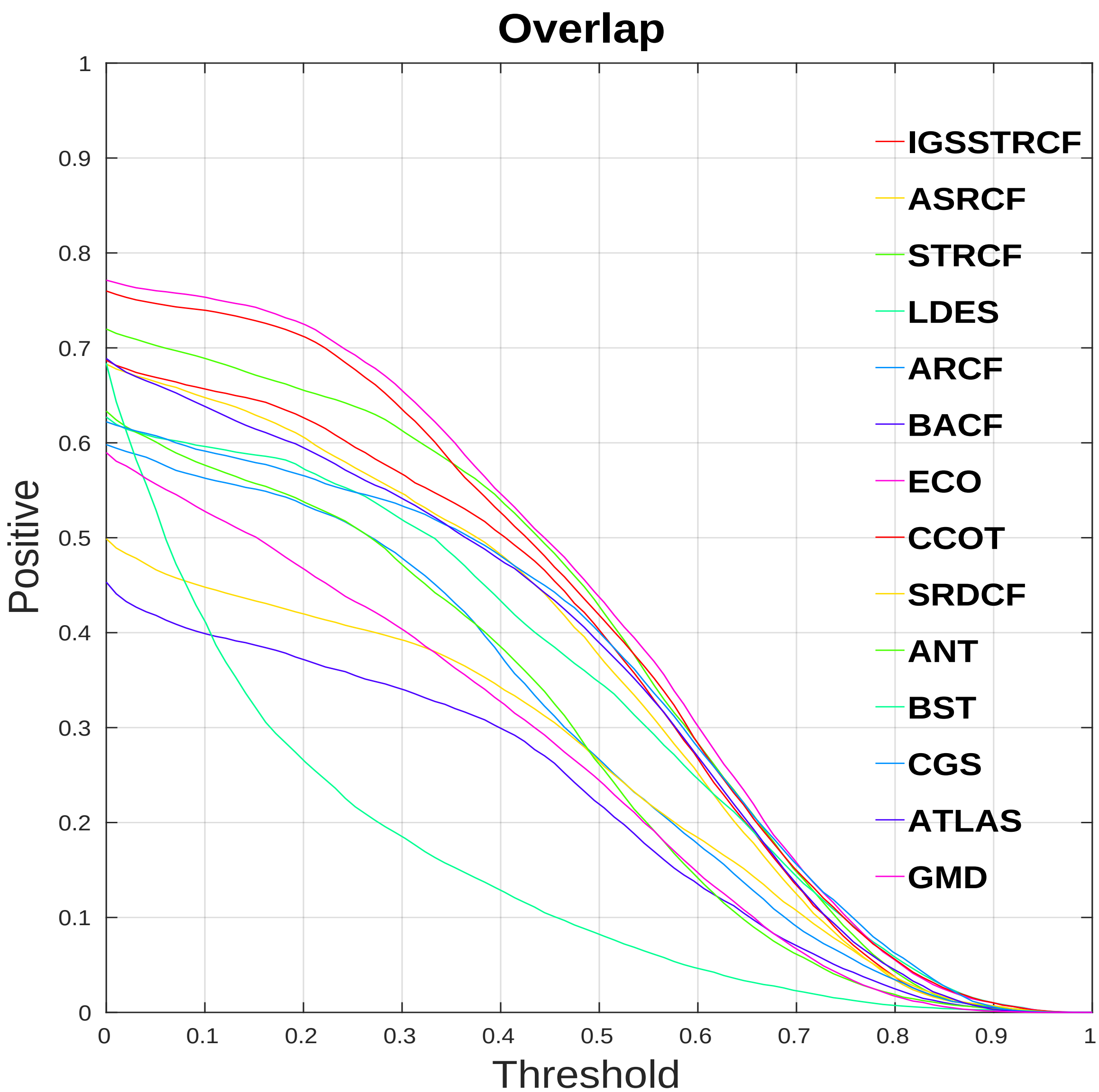

Figure 8: Expected Overlap curves for unsupervised experiments on VOT-2017 [7]

Table 4: Accuracy comparison of the proposed IGSSTRCF approach with recent trackers for the baseline experiments on VOT-2017 [7]. Here CM = Camera Motion, EMP = Empty Tag, IV= Illumination Variation, MC = Motion Change, OCC = Occlusion, SC = Size Change, and Wt. Mean = Weighted Mean. The top three trackers are shown in **red**, **blue** and **green**

	CM	**EMP**	**IV**	**MC**	**OCC**	**SC**	**Mean**	**Wt. Mean**	**Pooled**
IGSSTRCF	0.5129	0.4711	0.4821	0.4666	0.4076	**0.4610**	0.4669	0.4761	0.4784
ASRCF [1]	0.4906	0.4725	0.4809	0.4479	0.4010	0.4506	0.4572	0.4654	0.4701
STRCF [9]	0.4910	0.4470	0.4750	0.4400	0.3690	0.4330	0.4430	0.4510	0.4550
LDES [10]	**0.5370**	0.4956	0.4783	**0.4850**	0.4053	**0.4559**	0.4762	**0.4929**	**0.5023**
ARCF [5]	0.4883	0.4633	0.4525	0.4551	0.4100	0.4393	0.4514	0.4615	0.4647
BACF [6]	0.4845	0.4515	0.4532	0.4638	**0.4339**	0.3440	0.4385	0.4476	0.4526
ECO [2]	0.5131	0.4846	**0.5026**	0.4810	0.3520	0.4451	0.4631	0.4762	0.4848
CCOT [4]	**0.5158**	0.4994	0.4460	**0.4835**	0.3917	0.4499	0.4644	**0.4851**	**0.4949**
SRDCF [3]	0.4816	**0.5080**	**0.5796**	0.4492	**0.4171**	0.4416	**0.4795**	0.4767	0.4867
ANT [7]	0.4890	0.4541	0.3986	0.4504	0.4099	0.4166	0.4364	0.4540	0.4622
BST [7]	0.2376	0.3091	0.2606	0.2391	0.2412	0.2443	0.2553	0.2627	0.2697
CGS [7]	**0.5380**	**0.5018**	0.4831	0.4820	**0.4289**	**0.4632**	**0.4828**	**0.4979**	**0.5063**
ATLAS [7]	0.4926	**0.5079**	**0.5248**	**0.4830**	0.4098	0.4449	**0.4772**	0.4835	0.4916
GMD [7]	0.4607	0.4573	0.4258	0.4411	0.3801	0.4072	0.4287	0.4422	0.4490

Table 5: Robustness comparison of the proposed IGSSTRCF approach with recent trackers for the baseline experiments on VOT-2017 [7], following the same abbreviations as Table 4. The top three trackers are shown in **red**, **blue** and **green**

	CM	**EMP**	**IV**	**MC**	**OCC**	**SC**	**Mean**	**Wt. Mean**	**Pooled**
IGSSTRCF	**44.00**	**16.00**	6.00	**25.00**	**21.00**	**20.00**	**22.00**	**26.28**	**93.00**
ASRCF [1]	45.00	26.00	6.00	29.00	**23.00**	23.00	25.33	30.97	105.00
STRCF [9]	85.00	68.00	9.00	46.00	39.00	31.00	46.16	61.83	198.00
LDES [10]	47.00	46.00	10.00	31.00	31.00	27.00	32.00	39.64	133.00
ARCF [5]	53.00	49.00	8.00	29.00	34.00	**20.00**	32.16	41.41	141.00
BACF [6]	77.00	61.00	6.00	43.00	31.00	33.00	41.83	55.77	189.00
ECO [2]	**25.00**	**14.00**	**4.00**	**18.00**	**22.00**	**9.00**	**15.3333**	**17.66**	**59.00**
CCOT [4]	**26.00**	**19.00**	6.00	**19.00**	**22.00**	**14.00**	**17.66**	**20.41**	**68.00**
SRDCF [3]	76.00	86.00	9.00	49.00	32.00	29.00	46.83	64.11	208.00
ANT [7]	64.00	26.00	8.00	45.00	27.00	30.00	33.33	40.15	135.00
BST [7]	74.73	66.93	**4.86**	41.93	24.66	26.73	39.97	55.50	188.60
CGS [7]	73.66	62.46	8.40	39.26	24.13	26.73	39.11	53.37	172.06
ATLAS [7]	60.00	30.00	**2.00**	35.00	24.00	22.00	28.83	37.42	127.00
GMD [7]	86.06	50.33	5.40	47.66	30.73	26.20	41.06	54.73	187.46

Table 6: Overlap AUC comparison of the proposed IGSSTRCF approach with recent trackers for the unsupervised experiments on VOT-2017 [7], following the same abbreviations as Table 4. The top three trackers are shown in **red**, **blue** and **green**

	CM	**EMP**	**IV**	**MC**	**OCC**	**SC**	**OverAll**
IGSSTRCF	0.3677	**0.3532**	0.3288	0.3609	**0.2617**	0.3582	**0.3539**
ASRCF [1]	0.3583	0.3419	0.3083	0.3452	0.2511	0.3401	0.3411
STRCF [9]	0.3040	0.3180	0.2970	0.2701	0.1882	0.3141	0.3000
LDES [10]	0.3302	0.3405	**0.3648**	0.2988	0.1985	0.2929	0.3237
ARCF [5]	0.3055	0.2735	0.2781	0.2836	0.1989	0.3046	0.2824
BACF [6]	0.2196	0.2204	0.2649	0.2021	0.1809	0.1472	0.2083
ECO [2]	**0.4189**	**0.4132**	**0.4251**	**0.3701**	**0.2804**	**0.3695**	**0.4025**
CCOT [4]	**0.4002**	**0.4113**	0.3618	0.3386	**0.2767**	0.3532	**0.3909**
SRDCF [3]	0.2606	0.2230	0.2901	0.2417	0.1990	0.2534	0.2445
ANT [7]	0.3036	0.2681	0.1910	0.2822	0.2047	0.2909	0.2770
BST [7]	0.1411	0.1678	0.1720	0.1067	0.1194	0.1391	0.1458
CGS [7]	**0.3773**	0.3014	0.3154	**0.3722**	0.2497	**0.3846**	0.3386
ATLAS [7]	0.3411	0.3524	**0.4237**	**0.3666**	0.2370	**0.3669**	0.3431
GMD [7]	0.2707	0.2486	0.1932	0.2262	0.1683	0.2886	0.2492

Table 7: Ablation analysis on VOT-2019 [8], VOT-2017 [7], TC128 [11] and UAV123 [12] to demonstrate comparison of the regularized version with its respective baselines. For VOT datasets, we report overlap score for the baseline experiments. For TC128 and UAV123, we report overlap success plot AUC and distance precision score at a threshold of 20 pixel. The best performing method is shown in red color

	VOT-2019	VOT-2017	TC128		UAV123	
	Overlap Score	Overlap Score	Success	Precision	Success	Precision
BACF [6]	0.116	0.123	48.45	63.81	40.01	45.20
BACF [6] + CI	0.157	0.178	57.50	78.29	52.38	67.44
STRCF [9]	0.114	0.118	50.64	67.58	47.88	61.79
STRCF [9] + STR	0.127	0.151	56.11	76.46	51.30	65.79
ASRCF [1]	0.145	0.185	57.14	76.90	47.06	60.96
ASRCF [1] + SSR	0.153	0.193	56.68	77.30	52.05	67.60

3. Ablation Study: Comparison with the Baselines

To demonstrate the performance of individual regularization terms independently, we add the regularization to the baseline tracker that is closest to the regularized version. The details are given as follows.

The Channel importance (CI) term is added to the baseline BACF [6]. Equation (1) shows the BACF [6] with the CI term (denoted 'BACF [6] + CI'). The black text in (1) is the original BACF [6] formulation.

$$E(\mathbf{h}) = \frac{1}{2}\left\| \mathbf{y} - \sum_{k=1}^{K} \overbrace{q_k}^{\textbf{CI}} (\mathbf{x}_k * (\mathbf{P}^T\mathbf{h}_k)) \right\|_2^2 + \frac{\lambda_1}{2}\sum_{k=1}^{K} \|\mathbf{h}_k\|_2^2 + \underbrace{\frac{\beta}{2}\|\mathbf{q}\|_2^2}_{\textbf{CI}}, \tag{1}$$

The Sparse Temporal Regularization (STR) is added to the baseline STRCF [9]. Equation (2) shows the STRCF [9] with the STR term (denoted 'STRCF [9] + STR'). The black text in (2) is the original STRCF [9] formulation.

$$E(\mathbf{h},\mathbf{w}) = \frac{1}{2}\left\| \mathbf{y} - \sum_{k=1}^{K} (\mathbf{x}_k * \mathbf{h}_k) \right\|_2^2 + \frac{\lambda_1}{2}\sum_{k=1}^{K} \|\mathbf{w} \odot \mathbf{h}_k\|_2^2 + \underbrace{\frac{\theta}{2}\left\| \mathbf{h}^{(t)} - \mathbf{h}^{(t-1)} - \mathbf{B}_h \right\|_2^2 + \eta \|\mathbf{B}_h\|_1}_{\textbf{STR}}, \tag{2}$$

Similarly, the Sparse Spatial Regularization (SSR) is added to the baseline ASRCF [1]. Equation (3) shows the ASRCF [1] with the SSR term (denoted 'ASRCF [1] + SSR'). The black text in (3) is the original ASRCF [1] formulation.

$$E(\mathbf{h},\mathbf{w}) = \frac{1}{2}\left\| \mathbf{y} - \sum_{k=1}^{K} (\mathbf{x}_k * (\mathbf{P}^T\mathbf{h}_k)) \right\|_2^2 + \frac{\lambda_1}{2}\sum_{k=1}^{K} \|\mathbf{w} \odot \mathbf{h}_k\|_2^2 + \underbrace{\frac{\lambda_2}{2}\left\| \mathbf{w} - \mathbf{w}_r - \mathbf{B}_w \right\|_2^2 + \zeta \|\mathbf{B}_w\|_1}_{\textbf{SSR}}, \tag{3}$$

Table 7 shows comparison of the regularized version with their respective baselines on VOT-2019 [8], VOT-2017 [7], TC128 [11] and UAV123 [12]. It is observed that adding the regularization results in improving the baseline performance in all the cases, except for success in TC128 dataset.

References

[1] Kenan Dai, Dong Wang, Huchuan Lu, Chong Sun, and Jianhua Li. Visual tracking via adaptive spatially-regularized correlation filters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4670–4679, 2019.

[2] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Eco: Efficient convolution operators for tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6638–6646, 2017.

[3] Martin Danelljan, Gustav Hager, Fahad Shahbaz Khan, and Michael Felsberg. Learning spatially regularized correlation filters for visual tracking. In *Proceedings of the IEEE international conference on computer vision*, pages 4310–4318, 2015.

[4] Martin Danelljan, Andreas Robinson, Fahad Shahbaz Khan, and Michael Felsberg. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision*, pages 472–488. Springer, 2016.

[5] Ziyuan Huang, Changhong Fu, Yiming Li, Fuling Lin, and Peng Lu. Learning aberrance repressed correlation filters for real-time uav tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2891–2900, 2019.

[6] Hamed Kiani Galoogahi, Ashton Fagg, and Simon Lucey. Learning background-aware correlation filters for visual tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1135–1143, 2017.

[7] Matej Kristan, Ales Leonardis, Jiri Matas, Michael Felsberg, et al. The visual object tracking vot2017 challenge results. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1949–1972, 2017.

[8] Matej Kristan, Jiri Matas, Ales Leonardis, Michael Felsberg, et al. The seventh visual object tracking vot2019 challenge results. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[9] Feng Li, Cheng Tian, Wangmeng Zuo, Lei Zhang, and Ming-Hsuan Yang. Learning spatial-temporal regularized correlation filters for visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4904–4913, 2018.

[10] Yang Li, Jianke Zhu, Steven CH Hoi, Wenjie Song, Zhefeng Wang, and Hantang Liu. Robust estimation of similarity transformation for visual object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8666–8673, 2019.

[11] Pengpeng Liang, Erik Blasch, and Haibin Ling. Encoding color information for visual tracking: Algorithms and benchmark. *IEEE Transactions on Image Processing*, 24(12):5630–5644, 2015.

[12] Matthias Mueller, Neil Smith, and Bernard Ghanem. A benchmark and simulator for uav tracking. In *European Conference on Computer Vision*, pages 445–461. Springer, 2016.