Computational Visual Media
https://doi.org/10.1007/

Review Article

Supplementary Material of Multi-modal Visual Tracking: Review and Experimental Comparison

Pengyu Zhang[1], **Dong Wang**(✉)[1], **and Huchuan Lu**[1]

Abstract

Keywords

Due to the page limitation, we describe the additional details in this supplementary material. In the section 1, we introduce the attribute annotation in RGB-D and RGB-T datasets and we depict the attirbute-based performance in RGBT234 dataset in section 2. Finally, we provide a detailed description of multi-modal tracking following the taxonomy in section 3.

1 Introduction of the attributes in STC and RGBT234 datasets

In this section, taking STC and RGBT234 as example, we detail the attribute annotation in RGB-D and RGB-T datasets. As for STC dataset, 10 types of attribute are annotated, including illumination variation, depth variation, scale variation, depth distribution variation, surrounding depth clutter, surrounding color clutter, background color camouflages, background shape camouflages and partial occlusion. RGBT234 dataset contains 12 attributes, including no occlusion, partial occlusion, heavy occlusion, low illumination, low resolution, thermal crossover, deformation, fast motion, scale variation, motion blur, camera moving and background clutter. The detailed description of attributes in both two datasets are shown in Table 1 and Table 2, respectively.

2 Attribute-based Comparison on RGBT234 dataset

Here, we give the detailed results with respect to attribute-based comparison on RGBT234 dataset, shown in Figure 1. Two mainstream frameworks, including deep learning based method and CF tracker obtain very competitive results according to attribute-based performance. Improved MDNet-based tracker, i.e. CMPP, shows superior performance in 6 attributes, such as low resolution, deformation, background clutter, thermal crossover, fast motion and heavy occlusion. JMMAC models both appearance and motion cues in a unified framework, thereby achieving satisfying performance in camera moving and partial occlusion. Furthermore, CF based trackers is more capable of handling scale variation than MDNet variants.

3 Detailed description of multi-modal trackers

We provide an in-depth description on multi-modal trackers. Following the taxonomy in the main paper, we summarize existing RGB-D and RGB-T tracking methods in Table 4 and Table 5 from various aspects, including, tracking framework, learning paradigm, feature type, auxiliary modality purpose. Furthermore, the public resources for multi-modal tracking are concluded in Table 3 for further research.

Table 1 Description of attributes annotated in the STC dataset.

Attribute	Description
IV	Illumination Variation – RGB intensity change of the target (mean value).
DV	Depth Variation – depth change of the target (mean value).
SV	Scale Variation – scale change of the bounding box (relative ratio).
CDV	Color Distribution Variation – RGB distribution change of the target.
DDV	Depth Distribution Variation – depth distribution change of the target.
SDC	Surrounding Depth Clutter – depth similarity between the target and contextual region (mean value).
SCC	Surrounding Color Clutter – RGB intensity similarity between the target and contextual region.
BCC	Background Color Camouflages – The distractor has a similar color as the target.
BSC	Background Shape Camouflages – The distractor has a similar shape as the target.
PO	Partial Occlusion – A part of the target is obstructed.

Table 2 Description of attributes annotated in RGBT210 and RGBT234 datasets.

Attribute	Description
NO	No Occlusion – The target is not occluded.
PO	Partial Occlusion – The target is partially occluded.
HO	Heavy Occlusion – over 80% of the target region is occluded.
LI	Low Illumination – The illumination in the target is low.
LR	Low Resolution – The resolution in the target is low.
TC	Thermal Crossover – The target has similar temperature with other objects or background.
DEF	Deformation – Non-rigid object deformation.
FM	Fast Motion – The target movement is larger than 20 pixels between two adjacent frames.
SV	Scale Variation – Compared with the initial target patch, the ratio of the target varies in a wide range.
MB	Motion Blur – The target object motion results in the blur image information.
CM	Camera Moving – The target object is captured by a moving camera.
BC	Background Clutter – The distractor has a similar color or shape as the target.

Table 3 Public resources for RGB-T and RGB-D tracking.

		Method	links
RGB-D	Tracker	3DT	https://github.com/adelbibi/3D-Part-Based-Sparse-Tracker-with-Automatic-Synchronization-and-Registration
		OTR	https://github.com/ugurkart/OTR
		DSOH	https://github.com/mcamplan/DSKCF_BMVC2015
		DSKCF	https://github.com/mcamplan/DSKCF_JRTIP2016
		ARDM	https://github.com/shine636363/RGBDtracker
		CSR-RGBD	http://tracking.cs.princeton.edu/
	Dataset	PTB	http://tracking.cs.princeton.edu/
		STC	https://beardatashare.bham.ac.uk/dl/fiVnhJRjkyNN8QjSAoiGSiBY/RGBDdataset.zip
		CDTB	https://www.vicos.si/Projects/CDTB
RGB-T	Tracker	MANet	https://github.com/Alexadlu/MANet
		mfDiMP	https://github.com/zhanglichao/end2end_rgbt_tracking
	Dataset	OTCBVS	http://vcipl-okstate.org/pbvs/bench/
		LITIV	https://www.polymtl.ca/litiv/en/codes-and-datasets
		GTOT	https://docs.google.com/uc?id=0B-Z6TyBF2ceIZ0c1anVhaHQ3MFk&export=download
		RGBT234	https://sites.google.com/view/ahutracking001/

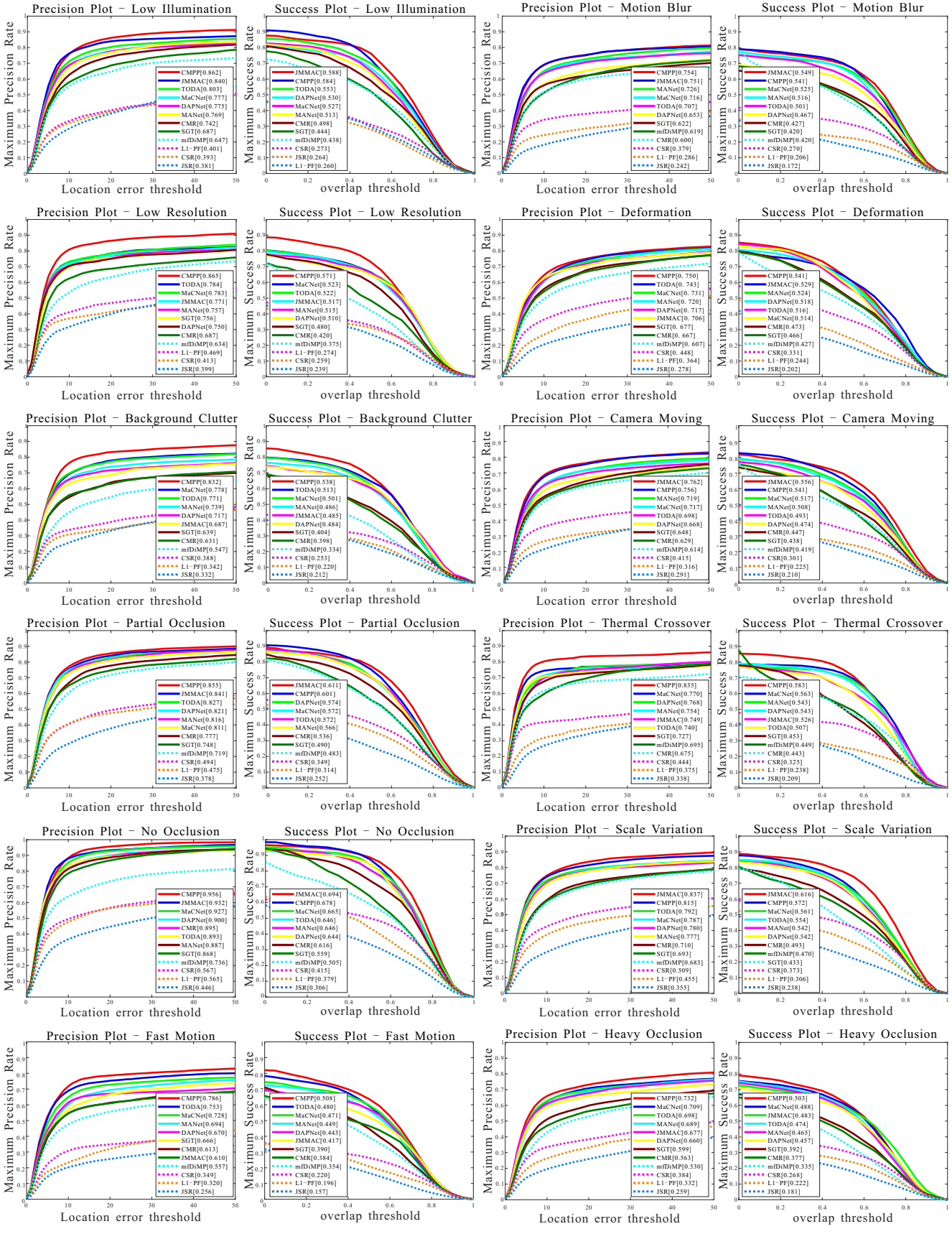

Fig. 1 Attribute-based Comparison on RGBT234.

Table 4 Detailed descriptions of RGB-D trackers.

	Year	Trackers	Framework	Learning Paradigm	Feature Type	Auxiliary Modality Purpose	Publication
RGB-D	2012	AMCT [1]	PF	Online learning	Scalable Gradient [2], Color Averages	Feature	JDOS
	2014	MCBT [3]	Dis. Others	Off-the-shelf	Optical Flow [4], Color Histograms [5]	Feature	Neu.
	2015	ISOD [6]	Dis. Others	Online learning	Intensity	Occlusion reasoning	SP
		DSOH [7]	CF	Online learning	HOG [8]	Feature, Scale estimation, Occlusion Reasoning	BMVC
		DOHR [9]	Dis. Others	Online learning	Haar-like [10]	Feature, Occlusion reasoning	FSKD
		CDG [11]	Dis. Others	Online learning	Depth Gradient, HOG, Depth Gradient Flow, Optical Flow [12]	Feature, Occlusion Reasoning	CAC
		OL3DC [13]	Dis. Others	Off-the-shelf	SURF [14]	3D Reconstruction, Occlusion Reasoning	Neu.
	2016	DLST [15]	CF	Online learning	HOG, Color Names [16]	Feature	ICPR
		DSKCF [17]	CF	Online learning	HOG	Feature, Occlusion Reasoning, Scale Estimation	RTIP
		3DT [18]	PF	Online learning	Color Names, 3D shape [19]	3D Reconstruction, Occlusion Reasoning	CVPR
		OAPF [20]	PF	Online learning	HOG, LBP, raw pixel, LoG, depth	Feature, Occlusion Reasoning, Scale Estimation	CVIU
	2017	ROTSL [21]	PF	Online learning	Depth-enhanced Color Feature	Feature, Occlusion Reasoning	ITEE
	2018	CSR-RGBD [22]	CF	Online learning	HOG, VGGNet-M [23]	Feature, Occlusion Reasoning	ECCVW
		DMDCF [24]	CF	Online learning	HOG, Color Names, Intensity	Feature, Occlusion Reasoning	ICPR
		SEOH [25]	CF	Online learning	HOG, Intensity	Feature, Scale Estimation	Access
		ARDM [26]	CF	Online learning	HOG, Color Names, Color Histogram	Feature	TC
		OACPF [27]	CF	Online learning	HOG	Occlusion Reasoning, Scale Estimation	Access
		CCF [28]	CF	Online learning	HOG, Color Names, Intensity	Occlusion Reasoning, Scale Estimation	GSKI
		RTKCF [29]	CF	Online learning	HOG, Color Names	Feature, Occlusion Reasoning	CCDC
	2019	3DMS [30]	MS	Online learning	Color Histogram	3D Reconstruction	ICST
		OTR [31]	CF	Online learning	HOG, Color Names	Feature, 3D Reconstruction	CVPR
		TACF [32]	CF	Online learning	HOG, Color Names, Lookup Table feature [33]	Feature, Occlusion Reasoning	Sensors
		CA3DMS [34]	MS	Online learning	Color Histogram	3D Reconstruction, Occlusion Reasoning	TMM
		OTOD [35]	DL	Online learning	PointNet [35]	3D Reconstruction	CIS
	2020	WCO [36]	CF	Online learning	HOG, VGGNet-M	Feature	Sensors

Table 5 Detailed description of RGB-T trackers.

	Year	Trackers	Framework	Learning Paradigm	Feature Type	Auxiliary Modality Purpose	Publication
RGB-T	2006	CFM [37]	Dis. Others	Online learning	Multi-dimensional Gaussian Feature	Feature	ICIF
	2007	PLF [38]	PF	Online learning	Colour Histogram	Feature	CVPR
	2008	MST [39]	MS	Online learning	Color Spatiograms [40]	Feature	MVA
		PGM [41]	Gen. Others	Online learning	Intensity	Feature	ISCS
	2011	JSR [42]	PF	Online learning	Color Histogram, Intensity	Feature	IS
		L1-PF [43]	PF	Online learning	Intensity, Canny Edge Description [44]	Feature	ICIF
	2016	RT-LSR [45]	SL	Online learning	Sparse feature	Feature	Multimedia
		CSR[46]	SL	Online learning	Intensity	Feature	TIP
	2017	SGT [47]	SL	Online learning	Sparse feature	Feature	Multimedia
		MLSR [48]	SL	Online learning	Sparse feature	Feature	TSMCS
	2018	RCDL [49]	SL	Online learning	HOG, Intensity	Feature	AAAI
		MSR [50]	SL	Online learning	Sparse feature	Feature	PRL
		CMR [51]	Dis. Others	Online learning	Raw pixel, intensity, HOG	Feature	ECCV
		RMR [52]	Dis. Others	Online learning	HOG, Color Histogram	Feature	SPIC
		LGMG [53]	SL	Online learning	HOG, Color Histogram	Feature	TCSVT
		MDNet-RGBT [54]	DL	Online learning, Offline training	VGGNet-M	Feature	CISP
		FTSNet [55]	CF	Online learning	VGGNet-16	Feature	Neu.
		CSCF [56]	CF	Online learning	Intensity	Feature	BICS
	2019	DAPNet [57]	DL	Online learning, Offline training	VGGNet-M	Feature	Multimedia
		HTF [58]	CF	Online learning	HOG, Color Names, Intensity	Feature	IPT
		LMCFT [59]	SL	Online learning	Sparse feature	Feature	TIE
		MANet [60]	DL	Online learning, Offline training	VGGNet-M	Feature	ICCVW
		TODA [61]	DL	Online learning, Offline training	VGGNet-M	Feature	ICIP
		DAFNet [62]	DL	Online learning, Offline training	VGGNet-M	Feature	ICCVW
		DiMP-RGBT [63]	DL	Online learning, Offline training	ResNet-101 [64]	Feature	ICCVW
		ONMF [65]	SL	Online learning	Sparse feature	Feature	Access
	2020	CMPP [66]	DL	Online learning, Offline training	VGGNet-M	Feature	CVPR
		MaCNet [67]	DL	Online learning, Offline training	VGGNet-M	Feature	Sensors
		CAT [68]	DL	Online learning, Offline training	VGGNet-M	Feature	ECCV
	2021	JMMAC [69]	CF	Online learning, Offline training	VGGNet-M	Feature, Occlusion Reasoning	TIP

References

[1] Garcia GM, Klein DA, Stuckler J. Adaptive Multi-cue 3D Tracking of Arbitrary Objects. In *Joint DAGM and OAGM Symposium*, 2012, 357–366.

[2] Klein DA, Cremers AB. Boosting Scalable Gradient Features for Adaptive Real-Time Tracking. In *International Conference on Robotics and Automation*, 2011, 4411–4416.

[3] Wang Q, Fang J, Yuan Y. Multi-cue Based Tracking. *Neurocomputing*, 2014, 131: 227–236.

[4] Brox T, Bruhn A, Papenberg N, Weickert J. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In *European Conference on Computer Vision*, 2004, 25–36.

[5] Porikli FM. Integral Histogram: a Fast Way to Extract Histograms in Cartesian Spaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, 829–836.

[6] Chen Y, Shen Y, Liu X, Zhong B. 3D object tracking via image sets and depth-based occlusion detection. *Signal Processing*, 2015, 112: 146–153.

[7] Camplani M, Hannuna S, Mirmehdi M, Damen D, Paiement A, Tao L, Burghardt T. Real-time RGB-D Tracking with Depth Scaling Kernelised Correlation Filters and Occlusion Handling. In *British Machine Vision Conference*, 2015, 145.1–145.11.

[8] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, 886–893.

[9] Ding P, Song Y. Robust Object Tracking using Color and Depth Images with a Depth based Occlusion Handling and Recovery. In *International Conference on Fuzzy Systems and Knowledge Discovery*, 2015, 930–935.

[10] Viola P, Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, 511–518.

[11] Shi H, Gao C, Sang N. Using Consistency of Depth Gradient to Improve Visual Tracking in RGB-D Sequences. In *Chinese Automation Congress*, 2015, 518–522.

[12] KPHorn B, GSchunck B. Determining Optical Flow. *Artificial Intelligence*, 1981, 17: 185–203.

[13] Zhong B, Shen Y, Chen Y, Xie W, Cui Z, Zhang H, Chen D, Wang T, Liu X, Peng S, Gou J, Du J, Wang J, Zheng W. Online Learning 3D Context for Robust Visual Tracking. *Neurocomputing*, 2015, 151: 710–718.

[14] Bay H, Ess A, Tuytelaars T, Gool LV. Speeded-Up Robust Features. *Computer Vision and Image Understanding*, 2008, 110: 346–359.

[15] An N, Zhao XG, Hou ZG. Online RGB-D Tracking via Detection-Learning-Segmentation. In *International Conference on Pattern Recognition*, 2016, 1231–1236.

[16] van de Weijer J, Schmid C, Schmid C, Larlus D. Learning Color Names for Real-World Applications. *IEEE Transactions on Image Processing*, 2009, 18(7): 1512–1523.

[17] Hannuna S, Camplani M, Hall J, Mirmehdi M, Damen D, Burghardt T, Paiement A, Tao L. DS-KCF: a real-time tracker for RGB-D data. *Journal of Real-Time Image Processing*, 2019, 16: 1439–1458.

[18] Bibi A, Zhang T, Ghanem B. 3D Part-Based Sparse Tracker with Automatic Synchronization and Registration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 1439–1448.

[19] Song S, Xiao J. Tracking Revisited using RGBD Camera: Unified Benchmark and Baselines. In *IEEE International Conference on Computer Vision*, 2013, 233–240.

[20] Meshgi K, ichi Maeda S, Oba S, Skibbe H, zhe Li Y, Ishii S. An Occlusion-aware Particle Filter Tracker to Handle Complex and Persistent Occlusions. *Computer Vision and Image Understanding*, 2016, 150: 81–94.

[21] Ma Z, Xiang Z. Robust Object Tracking with RGBD-based Sparse Learning. *Frontiers of Information Technology and Electronic Engineering*, 2017, 18(7): 989–1001.

[22] Kart U, Kamarainen JK, Matas J. How to Make an RGBD Tracker? In *European Conference on Computer Vision Workshop*, 2018, 148–161.

[23] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*, 2015, 1–14.

[24] Kart U, Kamarainen JK, Matas J, Fan L, Cricri F. Depth Masked Discriminative Correlation Filter. In *International Conference on Pattern Recognition*, 2018, 2112–2117.

[25] Leng J, Liu Y. Real-time RGB-D Visual Tracking with Scale Estimation and Occlusion Handling. *Access*, 2018, 6: 24256–24263.

[26] Xiao J, Stolkin R, Gao Y, Leonardis A. Robust Fusion of Color and Depth Data for RGB-D Target Tracking Using Adaptive Range-Invariant Depth Models and Spatio-Temporal Consistency Constraints. *IEEE Transactions on Cybernetics*, 2018, 48(8): 2485–2499.

[27] Zhai Y, Song P, Mou Z, Chen X, Liu X. Occlusion-Aware Correlation Particle Filter Target Tracking Based on RGBD Data. *Access*, 2018, 6: 50752–50764.

[28] Li G, Huang L, Zhang P, Li Q, Huo Y. Depth Information Aided Constrained Correlation Filter for Visual Tracking. In *International Conference on Geo-Spatial Knowledge and Intelligence*, 2019, 1–10.

[29] Zhang H, Cai M, Li J. A Real-time RGB-D tracker based on KCF. In *Chinese Control And Decision Conference*, 2018, 4856–4861.

[30] Gutev A, Debono CJ. Exploiting Depth information to increase object tracking robustness. In *International Conference on Smart Technologies*, 2019, 1–5.

[31] Kart U, Lukezic A, Kristan M, Kamarainen JK, Matas J. Object Tracking by Reconstruction with View-Specific Discriminative Correlation Filters. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, 1339–1348.

[32] Kuai Y, Wen G, Li D, Xiao J. Target-Aware Correlation Filter Tracking in RGBD Videos. *Sensors*, 2019, 19(20): 9522–9531.

[33] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, 2012, 1106–1114.

[34] Liu Y, Jing XY, Nie J, Gao H, Liu J. Context-Aware Three-Dimensional Mean-Shift With Occlusion Handling for Robust Object Tracking in RGB-D Videos. *IEEE Transactions on Multimedia*, 2019, 21(3): 664–677.

[35] Xie Y, Lu Y, Gu S. RGB-D Object Tracking with Occlusion Detection. In *International Conference on Computational Intelligence and Security*, 2019, 11–15.

[36] Liu W, Tang X, Zhao C. Robust RGBD Tracking via Weighted Convolution Operators. *Sensors*, 2020, 20(8): 4496–4503.

[37] Conaire CO, O'Connor NE, Cooke E, Smeaton AF. Comparison of Fusion Methods for Thermo-visual Surveillance Tracking. In *International Conference on Information Fusion*, 2006, 1–7.

[38] Cvejic N, Nikolov SG, Knowles HD, Loza A, Achim A, Achim A, Canagarajah CN. The Effect of Pixel-Level Fusion on Object Tracking in Multi-Sensor Surveillance Video. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, 1–7.

[39] Conaire CO, O'Connor NE, Smeaton AF. Thermo-visual Feature Fusion for Object Tracking using Multiple Spatiogram Trackers. *Machine Vision and Applications*, 2008, 19(5): 483–494.

[40] Birchfield S, Rangarajan S. Spatiograms Versus Histograms for Region-Based Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, 1158–1163.

[41] Chen S, Zhu W, Leung H. Thermo-Visual Video Fusion Using Probabilistic Graphical Model for Human Tracking. In *International Symposium on Circuits and Systems*, 2008, 1926–1929.

[42] Liu H, Sun F. Fusion tracking in color and infrared images using joint sparse representation. *Information Sciences*, 2012, 55(3): 590–599.

[43] Wu Y, Blasch E, Chen G, Bai L, Ling H. Multiple source data fusion via sparse representation for robust visual tracking. In *International Conference on Information Fusion*, 2011, 1–8.

[44] Canny J. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8(6): 679–698.

[45] Li C, Hu S, Gao S, , Tang J. Real-Time Grayscale-Thermal Tracking via Laplacian Sparse Representation. In *ACM International Conference on Multimedia*, 2016, 54–65.

[46] Li C, Cheng H, Hu S, Liu X, Tang J, Lin L. Learning Collaborative Sparse Representation for Grayscale-thermal Tracking. *IEEE Transactions on Image Processing*, 2016, 25(12): 5743–5756.

[47] Li C, Zhao N, Lu Y, Zhu C, Tang J. Weighted Sparse Representation Regularized Graph Learning for RGB-T Object Tracking. In *ACM International Conference on Multimedia*, 2017, 1856–1864.

[48] Li C, Sun X, Wang X, Zhang L, Tang J. Grayscale-Thermal Object Tracking via Multitask Laplacian Sparse Representation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 47(4): 673–681.

[49] Lan X, Ye M, Zhang S, Yuen PC. Robust Collaborative Discriminative Learning for RGB-Infrared Tracking. In *AAAI Conference on Artificial Intelligence*, 2018, 7008–7015.

[50] Lan X, Ye M, Zhang S, Zhou H, Yuen PC. Modality-correlation-aware Sparse Representation for RGB-infrared Object Tracking. *Pattern Recognition Letters*, 2018: 12–20.

[51] Li C, Zhu C, Huang Y, Tang J, Wang L. Cross-Modal Ranking with Soft Consistency and Noisy Labels for Robust RGB-T Tracking. In *European Conference on Computer Vision*, 2018, 831–847.

[52] Li C, Zhu C, Zheng S, Luo B, Tang J. Two-stage Modality-graphs Regularized Manifold Ranking for RGB-T Tracking. *Signal Processing: Image Communication*, 2018, 68: 207–217.

[53] Li C, Zhu C, Zhang J, Luo B, Wu X, Tang J. Learning Local-Global Multi-Graph Descriptors for RGB-T Object Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 29(10): 2913–2926.

[54] Zhang X, Zhang X, Du X, Zhou X, Yin J. Learning Multi-domain Convolutional Network for RGB-T Visual Tracking. In *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2018, 1–6.

[55] Li C, Wu X, Zhao N, Cao X, Tang J. Fusing two-stream convolutional neural networks for RGB-T object tracking. *Neurocomputing*, 2018, 281: 78–85.

[56] Wang Y, Li C, Tang J, Sun D. Learning Collaborative Sparse Correlation Filter for Real-Time Multispectral Object Tracking. In *International Conference on Brain Inspired Cognitive Systems*, 2018, 462–472.

[57] Zhu Y, Li C, Luo B, Tang J, Wang X. Dense Feature Aggregation and Pruning for RGBT Tracking. In *ACM International Conference on Multimedia*, 2019, 465–472.

[58] Luo C, Sun B, Yang K, Lu T, Yeh WC. Thermal Infrared and Visible Sequences Fusion Tracking based on a Hybrid Tracking Framework with Adaptive Weighting Scheme. *Infrared Physics and Technology*, 2019, 99: 265–276.

[59] Lan X, Ye M, Shao R, Zhong B, Yuen PC, Zhou H. Learning Modality-Consistency Feature Templates: A Robust RGB-Infrared Tracking System. *IEEE Transactions on Industrial Electronics*, 2019, 66(12): 9887–9897.

[60] Li C, Lu A, Zheng A, Tu Z, Tang J. Multi-Adapter RGBT Tracking. In *IEEE International Conference on Computer Vision Workshop*, 2019, 2262–2270.

[61] Yang R, Zhu Y, Wang X, Li C, Tang J. Learning Target-Oriented Dual Attention For Robust RGB-T Tracking. In *IEEE International Conference on Image Processing*, 2019, 3975–3979.

[62] Gao Y, Li C, Zhu Y, Tang J, He T, Wang F. Deep Adaptive Fusion Network for High Performance RGBT Tracking. In *IEEE International Conference on Computer Vision Workshop*, 2019, 91–99.

[63] Zhang L, Danelljan M, Gonzalez-Garcia1 A, van de Weijer J,

Khan FS. Multi-Modal Fusion for End-to-End RGB-T Tracking. In *IEEE International Conference on Computer Vision Workshop*, 2019, 2252–2261.

[64] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 770–778.

[65] Lan X, Ye M, Shao R, Zhong B. Online Non-Negative Multi-Modality Feature Template Learning for RGB-Assisted Infrared Tracking. *Access*, 2019, 7: 67761–67771.

[66] Wang C, Xu C, Cui Z, Zhou L, Zhang T, Zhang X, Yang J. Cross-Modal Pattern-Propagation for RGB-T Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, 7062–7071.

[67] Zhang H, Zhang L, Zhuo L, Zhang J. Object Tracking in RGB-T Videos Using Modal-Aware Attention Network and Competitive Learning. *Sensors*, 2020, 20(2).

[68] Li C, Liu L, Lu A, Ji Q, Tang J. Challenge-Aware RGBT Tracking. In *European Conference on Computer Vision*, 2020, 222–237.

[69] Zhang P, Zhao J, Bo C, Wang D, Lu H, Yang X. Jointly Modeling Motion and Appearance Cues for Robust RGB-T Tracking. *IEEE Transactions on Image Processing*, 2021, 30: 3335–3347.