MOSAIC SPATIAL-SPECTRAL FEATURE BASED OBJECT TRACKING IN HYPERSPECTRAL VIDEO

Lulu Chen[1,2], Yongqiang Zhao[1,2], *Member, IEEE*

[1] Research & Development Institute of Northwestern Polytechnical University in Shenzhen
[2]Northwestern Polytechnical University, School of Automation

ABSTRACT

Recent studies have shown that the problem of color trackers under challenging situations can be alleviate by the material information of hyperspectral image (HSI), which can be acquired at video rate by a snapshot mosaic hyperspectral camera with spectral filter array (SFA). Due to the specific mosaic structure of the acquired images, it is usually converted into an HSI cube directly or by demosaicing, which will reduce spatial resolution and cause spatial-spectral distortion. To this end, we propose a novel mosaic spatial-spectral tracking (MSST) framework for snapshot mosaic HS videos. First, considering the spatial-spectral correlation of mosaic HSI, the novel mosaic spatial and spectral gradient operators are designed dedicated to raw mosaic HSI. Then, mosaic spatial-spectral histogram of oriented gradient (MSSHOG) descriptor is constructed by exploring the distribution of gradient magnitudes in spatial domain and spatial-spectral domain. Finally, MSSHOG is further embedded to correlation filters, yielding MSST method. The experimental results demonstrate the feasibility and effectiveness of MSST.

***Index Terms*—**Snapshot mosaic hyperspectral camera, Mosaic spatial-spectral gradient operators, mosaic spatial-spectral histogram of oriented gradient, Spectral filter array, Mosaic spatial spectral tracking.

1. INTRODUCTION

Traditional video tracking often fails in challenging situations, e.g. deformation, background clutter and object rotation [1]. Hyperspectral image (HSI) [2-3] can characterize target with great precision and detail, which can handle with the object drift problem caused by the above challenges. The early work on HS tracking only used spectral information as features [4-6]. Uzkent et al. [7] extract deep feature by converting HSI to false-color image. Qian et al. [8] extract the feature using the 3D patches selected from object area in the first frame. However, these works fail to make full use of spatial-spectral information of HSI, and there is no large scale dataset to evaluate tracking performance.

Currently, Xiong et al. [9] proposed a material-based HS tracking (MHT) method and introduced a HS tracking dataset acquired by a snapshot mosaic HS camera equipped with 4×4 spectral filter arrays (SFA) [11-13]. And in [10] further proposed a dynamic material-aware tracking (DMT) method. MHT and DMT directly convert the raw mosaic HSI into a HS cube to extract material information. However, this conversion will reduces the spatial resolution and causes spatial distortion. Demosaicing can recover the fully-defined HS cube without reducing spatial resolution [14]. However, there is a deviation between the estimated values and the actual values, which will causes the spatial-spectral distortion. In addition, high dimension of fully-defined HS cube will bring high computational costs [15].

To address the above problem, we propose to extract feature directly from raw mosaic HSI, which can avoids demosaicing step and preserves spatial resolution. To our best knowledge, there is the first time to study the object tracking directly in snapshot mosaic HS video. According to the structure of SFA pattern, each filter is only sensitive to one spectral band, resulting in neighboring pixels associated with same band not corresponding to neighboring pixels in the actual scene. That is, the neighboring pixels in mosaic HSI have the reduced spatial-spectral correlations [16, 17]. Therefore, the unique SFA structure should be considered to explore the spatial-spectral information of mosaic HSI to realize the tracking.

In this paper, we develop a novel mosaic spatial-spectral tracking framework for object tracking in mosaic HS videos. Considering the characteristics of SFA, the novel mosaic spatial gradient and spectral gradient operators are designed to construct the mosaic spatial-spectral histograms of oriented gradient (MSSHOG) descriptor to directly extract spatial-spectral feature from raw mosaic HSI. The gradient operators are based on SFA structure of mosaic HS sensor rather than the acquired data, so the proposed descriptor is applicable to any SFA pattern. The MSSHOG are further integrated into correlation filters, yielding the MSST method.

This work was supported by the Science, Technology and Innovation Commission of Shenzhen Municipality under Grants JCYJ20170815162956949 and JCYJ20180306171146740, Key R & D plan of Shaanxi Province 2020ZDLGY07-11, and by the National Natural Science Foundation of China (NSFC) under Grant 61771391 and by natural science basic research plan in Shaanxi Province of China 2018JM6056.

Extensive experiments on mosaic HS videos show that MSST exhibits a better performance than the advanced trackers.

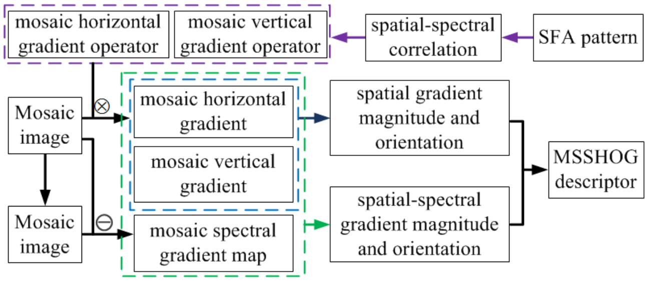
mosaic horizontal gradient operator
mosaic vertical gradient operator
spatial-spectral correlation
SFA pattern
Mosaic image
⊗
mosaic horizontal gradient
mosaic vertical gradient
spatial gradient magnitude and orientation
MSSHOG descriptor
Mosaic image
⊖
mosaic spectral gradient map
spatial-spectral gradient magnitude and orientation

Fig.1 The construction flowchart of MSSHOG descriptor.

2. METHODOLOGY

This section describe the details of MSST method.

2.1. Mosaic Spatial-spectral Histogram of Oriented Gradients

Considering the spatial-spectral aliasing information among the pixels of SFA pattern, we build MSSHOG descriptor to directly extract spatial-spectral feature from mosaic HSI. Fig. 1 shows the construction flowchart of MSSHOG descriptor. Given a mosaic HSI $I_M \in R^{X \times Y}$ containing $X \times Y$ pixels and K bands, the construction of MSSHOG descriptor is described as follows.

Mosaic spatial gradient: To computation the mosaic spatial gradient dedicated to the raw mosaic HSI, we first design the mosaic spatial gradient operators including mosaic vertical and horizontal gradient operators by analyzing the spatial-spectral correlation in local neighborhood of mosaic image. For simplicity, a regular grid both horizontally and vertically is used to obtain the spatial and spectral correlation. The spatial correlation is computed by the Euclidean distance for each pair of filters f_i and f_j , that is, $d_{ij} = d\left(f_i - f_j\right)$ for $i, j = 1,...,K$ ($K = k_1 \times k_2$ for $k_1 \times k_2$ mosaic pattern). Note that all pixels in one mosaic HSI have the same spatial correlation. Then, spatial correlation-based horizontal and vertical gradient operator f_{sx}^d and f_{sy}^d are:

$$f_{sx}^d = f_{sx}^8 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{1a}$$

$$f_{sy}^d = f_{sy}^8 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{1b}$$

where d is neighborhood size, here, d = 8 for illustration purposes.

Similarly, spectral correlation is represented by spectral distance, which is calculated by the absolute difference of center frequencies between two bands l and k, that is $d_{l-k} = \left| l_l - l_k \right|$. For $k_1 \times k_2$ SFA pattern, the SFA can be defined as $SFA = \{\lambda_1, \lambda_2, ..., \lambda_K\}, K = k_1 \times k_2$ (sort by spatial location of filters from left to right and top to bottom). l_i denotes that center frequencies of band associated with i position in SFA. For the pixel p associate with band λ_k , the set of bands N^k that are associated with the neighborhood of pixel p can be represented as $\{\lambda_{k-4}, \lambda_{k-3}, \lambda_{k-2}, \lambda_{k-1}, \lambda_{k+1}, \lambda_{k+1}, \lambda_{k+1}, \lambda_{k+1}\}$. Note that the neighborhood of p is always associated with the same N^k whatever the location of the pixel p associated with the band λ_k . So corresponding spectral correlation matrix f_k^d can be calculated below for pixel p associated with band λ_k .

$$f_k^d = f_k^8 \begin{bmatrix} 1/d_{\lambda_{k-4}-k} & 1/d_{\lambda_{k-3}-k} & 1/d_{\lambda_{k-2}-k} \\ 1/d_{\lambda_{k-1}-k} & 1 & 1/d_{\lambda_{k+1}-k} \\ 1/d_{\lambda_{k+2}-k} & 1/d_{\lambda_{k+3}-k} & 1/d_{\lambda_{k+4}-k} \end{bmatrix} \tag{2}$$

where $d_{\lambda_i - k}$ is the spectral distance between the bands pair $\lambda_i \in N^k$ and λ_k .

Based on above, the mosaic horizontal and vertical gradient operators f_{sx}^d and f_{sy}^d can be represented as:

$$f_{xk}^d = f_{sx}^d \otimes f_k^d \tag{3a}$$

$$f_{yk}^d = f_{sy}^d \otimes f_k^d \tag{3b}$$

Then the final mosaic spatial horizontal gradient G_x and vertical gradient G_y can be obtained by summarizing the results of convolving the raw image I_M with the f_{xk} , f_{yk} .

$$G_x = \underset{k=1}{\overset{K}{U}} I_M \otimes f_{xk} \tag{4a}$$

$$G_y = \underset{k=1}{\overset{K}{U}} I_M \otimes f_{yk} \tag{4b}$$

Mosaic spectral gradient: The pseudo-panchromatic image (PPI) denotes the average image over all channels of HSI and is strongly correlated with all channels [14]. The property allows us to represent the mosaic spectral gradient through the difference between mosaic HSI and PPI. Based on this, we designed the mosaic spectral gradient operators to directly compute the spectral gradient from mosaic HSI. Here, the PPI is estimated by an averaging filter M, which is calculated by spatial distance between the central filter and other filters in the neighborhood. Take 4×4 SFA as an example to represent M.

$$M = \frac{1}{81} \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 4 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix} \tag{5}$$

The mosaic spectral gradient can be obtained by the following calculation:

$$G_l = I_M - I_M \otimes M = I_M \otimes ([1] - M) = I_M \otimes f_l \qquad (6)$$

where f_l denotes mosaic spectral gradient operator, and [1] represents a 5×5 square matrix with a center pixel of 1, and zero elsewhere.

MSSHOG descriptor: After gradient computation, two mosaic spatial-spectral descriptors are constructed upon the spatial dimension and spatial-spectral dimension using the calculated gradient maps. For first descriptor, the gradient magnitude M_{xy} and angle orientation $\theta(x,y)$ are calculated by the two mosaic spatial gradient maps. For second descriptor, the gradient magnitude M_{xyl} and angle orientation $\phi(x,y,l)$ are calculated by two mosaic spatial gradient maps and one mosaic spectral gradient map. The corresponding gradient magnitudes and orientations are computed as follows.

$$M_{xy} = \sqrt{G_x^2 + G_y^2}\ ,\ \ \theta(x,y) = \tan^{-1}\left(G_x / G_y\right) \qquad (7)$$

$$M_{xyl} = \sqrt{G_x^2 + G_y^2 + G_l^2}\ ,\ \ \phi(x,y,l) = \tan^{-1}\left(G_l \middle/ \sqrt{G_x^2 + G_y^2}\right) \qquad (8)$$

Here, 9 sensitive directions and 18 insensitive directions within 360 degrees are used to created orientation maps [18]. Then histograms are calculated for each pair of magnitude and orientation. Finally, MSSHOG descriptor is obtained by connecting the two histograms in the third dimension in series. Consequently, vector feature of (9+18+4) ×2 dimensions is obtained.

2.2 Mosaic Spatial-Spectral Object Tracking

Our MSST is based on Spatial-Temporal Regularized Correlation Filters (STRCF) [19], where the spatial and temporal regularization is used to alleviate the boundary effect. Mathematically, the loss function is expressed as follows:

$$\min_f \frac{1}{2}\left\| y - \sum_{k=1}^{K} f_k \otimes h_k^t \right\|_2^2 + \frac{1}{2}\sum_{k=1}^{K}\left\| w \bullet h_k \right\|_2^2 + \frac{1}{2}\left\| h - h^{t-1} \right\|_2^2 \qquad (9)$$

Where K is the feature dimension, y is the desired response map and h^t is the correlation filter to be learned at t frame. f is the MSSHOG feature map. h follow the solution in [19].

3. EXPERIMENTS

In this section, we evaluate the effectiveness of the proposed MSST tracker relative to advanced trackers.

3.1. Experiment Setting

Dataset: In the experiments, we use the mosaic HS video tracking dataset provided by Xiong et al. [9], which is acquired using a snapshot mosaic HS camera equipped with 4×4 SFA. The whole dataset contains 35 fully-annotated sequences, which are labelled with associated challenge attributes, such as illumination variation (IV), scale variation (SV), occlusion (OCC), deformation (DEF), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out-of-view (OV), background clutter (BC) and low resolution (LR).

Evaluation Metrics: To describe the evaluation performance, this paper use precision plot, success plot, distance precision (DP), and overlap precision (OP) [20].

3.2. Evaluation comparison with color trackers

This experiments compare our MSST with 9 state-of-the-art color trackers, including five deep learning based trackers STRCF [19], GFSDCF [21], CFWCR [22], ECO [23], UDT [24] and MCCT [25], and three hand-crafted feature based trackers CACF [1], KCF [26], Autotrack [27]. For fair comparison, color trackers and our tracker are performed on the raw mosaic HS videos. Fig. 2 reports the precision and success plots of all trackers. Overall, our MSST outperforms all comparison trackers on both metrics. Compared with the second ranked tracker MCCT, our MSST achieves the improvements by 2.8% and 4.0% in precision rate and success rate due to the full use of the mosaic spatial-spectral information. This also implies that deep feature and hand-crafted feature based color trackers may fails to exploit spatial-spectral information of mosaic image since the color image structure is different with mosaic image. It is worth mentioning that the MSSHOG based MSST obtains a significant improvement over the original baseline STRCF with CNN feature, and provides a gain of 3.7% and 5.1% in average precision rate and success rate.

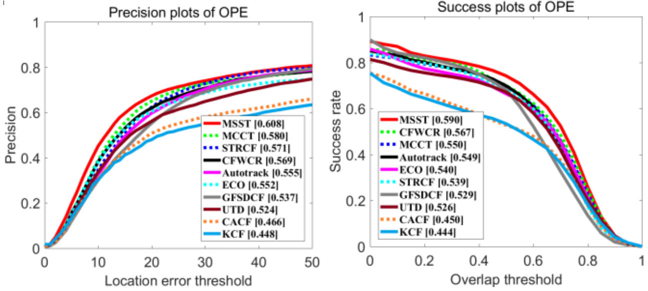

Fig. 2 Comparison with color trackers on mosaic videos.

3.3. Evaluation comparison with hyperspectral trackers

This section compare our MSST with two hyperspectral trackers, DeepHKCF [7] and MHT [9]. All HS trackers are run on the estimated fully-defined HS cube video obtained by performing demosaicing on raw mosaic HS video. Here we use WB method to demosaic since it is the most generic method. Fig.3 shows the comparison on precision and success plot. It shows that DeepHKCF give inferior accuracy, mainly because it converts HSI into a three-channel false-color image. Compared with MHT, our method performs better performance since it considers the spatial-spectral aliasing correlation to directly extract feature from raw mosaic HSI. The reason for the low performance of MHT may be the spatial-spectral distortion caused by demosaicing process. We also show the comparison in DP and OP in Table II, which shows the same results with Fig.3.

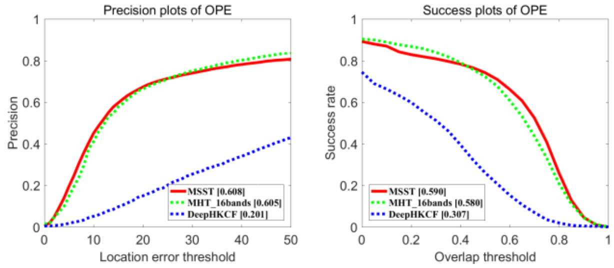

Fig. 3 Comparison with hyperspectral trackers.

3.4. Attribute based comparison

Table II reports the attribute based comparison results of all trackers. Here only presents the performance of top seven trackers mentioned above. We can find that our MSST is more competitive than the other trackers for handling challenging issues. It ranks the first on 6 out of 11 attributes. Compared with the second-ranked MHT tracker, MSST performs better in BC, IPR, OPR, MB, LR, and OV, which further demonstrates our MSST can not only enhance the robustness of spatial-spectral features, but also avoid the adverse effects of demosaicing. Additionally, the improvement of our method is more obvious compared with color trackers, which is due to the fact that the color trackers fails to exploit spatial-spectral aliasing information of mosaic HSI. In addition, it outperform STRCF on most attributes, which further show the effectiveness of our method and the feasibility and necessity of obtaining feature information directly from the raw mosaic HSI. In summary, our method has obvious advantages in handling various challenges problems over other features.

Table I Comparison With Hyperspectral Trackers in DP and OP.

	MSST	MHT	DeepHKCF
DP	0.674	0.663	0.150
OP	0.742	0.723	261

Table II Attribute-Based Comparison in Average Success Rate

	MSST	MHT[9]	ECO[23]	MCCT[25]	STRCF[19]	CFWCR[22]	GFSDCF[21]	Autotrack[27]
SV	0.569	0.578	0.501	0.567	0.523	0.556	0.494	0.554
MB	0.684	0.556	0.596	0.574	0.586	0.540	0.524	0.558
OCC	0.556	0.576	0.530	0.513	0.505	0.542	0.499	0.524
FM	0.530	0.536	0.422	0.593	0.565	0.560	0.510	0.572
LR	0.493	0.442	0.397	0.485	0.371	0.438	0.394	0.463
IPR	0.697	0.667	0.620	0.685	0.692	0.681	0.622	0.659
OPR	0.699	0.679	0.629	0.660	0.652	0.677	0.626	0.642
DEF	0.629	0.628	0.618	0.630	0.640	0.648	0.602	0.646
BC	0.646	0.620	0.595	0.538	0.561	0.579	0.613	0.527
IV	0.486	0.505	0.433	0.478	0.376	0.514	0.375	0.492
OV	0.661	0.624	0.638	0.592	0.485	0.591	0.364	0.649

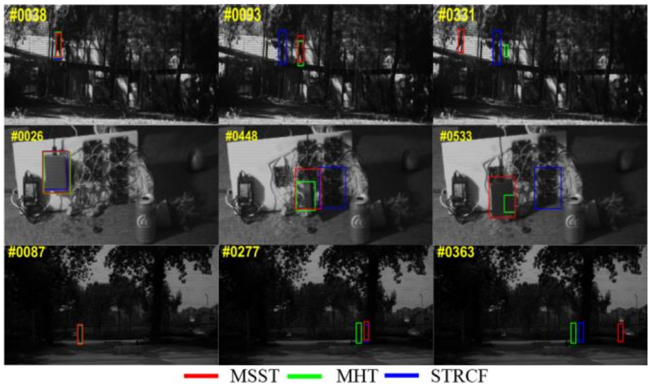

Fig. 4. Qualitative evaluation on three video sequences (i.e., campus, drive, and pedestrian2).

3.5. Visual Comparison

Fig.4 shows the qualitative evaluation of MSST, MHT and STRCF on three representative videos, in which there are background clutter, rotation, illumination various, low resolution. From Fig. 4, it can see that our MSST performs well in the whole sequence whereas other trackers have some deviations in scale and position or make the object drift to background. Overall, our MSST have the ability to handle with some challenges in tracking.

4. CONCLUSION

This paper presents a novel generic novel mosaic spatial-spectral tracking (MSST) framework that can be used for any SFA pattern. In MSST, based on SFA structure, the mosaic spatial and spectral gradient operators are designed to directly extract the spatial-spectral feature from raw mosaic

HSI. Then, by exploring the distribution of gradient magnitudes in spatial domain and spatial-spectral domain, respectively, a mosaic spatial-spectral histogram of oriented gradient (MSSHOG) descriptor is constructed. Finally, MSSHOG is further embedded to correlation filters, yielding the MSST method. Experimental results on mosaic HS video dataset show that the effectiveness of our method and the feasibility and necessity of extracting feature information directly from the raw mosaic HSI.

5. REFERENCES

[1] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017.

[2] C. Zhang, C. Guo, F. Liu, *et al.*, "Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine", *Journal of food engineering*, vol.17, no.9, pp.11-18, 2016.

[3] H. Lee, H. Kwon, "Going Deeper With Contextual CNN for Hyperspectral Image Classifi-cation", *IEEE Trans. Image Process.*, vol. 26, no. 10, pp.4843-4855, 2017.

[4] B. Uzkent, M. J. Hoffman and A. Vodacek, "Integrating hyperspectral likelihoods in a multidimensional assignment algorithm for aerial vehicle tracking", *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4325-4333, 2016.

[5] G. Tochon, J. Chanussot, M. Dalla Mura, and *et al.*, "Object tracking by hierarchical decomposition of hyperspectral video sequences: Application to chemical gas plume tracking", *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4567-4585, 2017.

[6] Y. Xu, Z. Wu, J. Chanussot, and *et al.*, "Low-rank decomposition and total variation regularization of hyperspectral video sequences", *IEEE Trans. Geosci. Remote Sens.,* vol. 56, no. 3, pp. 1680-1694, 2018.

[7] B. Uzkent, A. Rangnekar and M. J. Hoffman, "Tracking in aerial hyperspectral videos using deep kernelized correlation filters", *IEEE Trans. Geosci. Remote Sens.*, pp.1-13, 2018.

[8] K. Qian, J. Zhou, F. Xiong and *et al.*, "Object tracking in hyperspectral videos with convolutional features and kernelized correlation filter", *Proc. International Conference on Smart Multimedia,* 2018.

[9] F. Xiong, J. Zhou, Y. Qian, "Material Based Object Tracking in Hyperspectral Videos", *IEEE Trans. Image Process.,* 2020.

[10] F. Xiong, J. Zhou, J. Chanussot, *et al.*, "Dynamic Material-Aware Object Tracking in Hyperspectral Videos", *Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS),* 2019.

[11] N. Hagen, M. W. Kudenov, "Review of snapshot spectral imaging technologies", *Optical Engineering,* vol. 52, no. 9, 2013.

[12] M. Sofiane, "Snapshot multispectral image demosaicing and classification", 2018.

[13] J. B. Thomas, P. J. Lapray, P. Gouton, and *et al.*, "Spectral characterization of a proto-type SFA camera for joint visible and NIR acquisition", *Sensors,* vol. 16, no.7, 2016.

[14] G. Tsagkatakis, M. Bloemen, B. Geelen, *et al.*, "Graph and Rank Regularized Matrix Recovery for Snapshot Spectral Image Demosaicing", *IEEE Trans Compu. Imaging,* 2018.

[15] S. Mihoubi, O. Losson, B. Mathon, and *et al.*, "Multispectral demosaicing using pseudo-panchromatic image", *IEEE Trans. Compu. Imaging,* vol. 3, no. 4, pp. 982–995, 2017.

[16] Y. Monno, S. Kikuchi, M. Tanaka, and *et al.*, "A Practical One-Shot Multispectral Imaging System Using a Single Image Sensor", *Image Processing,* vol. 24, no. 10, pp. 3048-3059, 2015.

[17] N. Hagen, M. W. Kudenov, "Review of snapshot spectral imaging technologies", *Optical Engineering,* vol. 52, no. 9, 2013.

[18] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2005.

[19] F. Li, C. Tian, W. Zuo, and *et al.*, "Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking", *CVPR*, 2018.

[20] H. K. Galoogahi, A. Fagg and S. Lucey, "Learning background aware correlation filters for visual tracking", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 21-26, 2017.

[21] T. Xu, Z. H. Feng, X. J. Wu, and *et al.*, "Joint Group Feature Selection and Discriminative Filter Learning for Robust Visual Object Tracking", *IEEE Interna. Conf. Comput. Vis.*, 2019.

[22] Z. He, Y. Fan, J. Zhuang, and *et al.*, "Correlation Filters with Weighted Convolution Responses", *IEEE Interna. Conf. Comput. Vis., Workshop*, 2017.

[23] M. Danelljan, G. Bhat, F. S. Khan, and *et al.*, "ECO: Efficient convolution operators for tracking", *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 6638–6646, 2017.

[24] N. Wang, Y. Song, C. Ma, and *et al.*, "Unsupervised Deep Tracking", *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020.

[25] Wang N , Zhou W , Tian Q , et al. "Multi-Cue Correlation Filters for Robust Visual Tracking", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018.

[26] J. F. Henriques, R. Caseiro, P. Martins, and *et al.*, "High-speed tracking with kernelized correlation filters", *IEEE Conf. Comput.Vis. Pattern Recognit.,* vol. 37, no. 3, pp. 583–596, 2014.

[27] Y. Li, C. Fu, F. Ding, *et al.* "AutoTrack: Towards High-Performance Visual Tracking for UAV with Automatic Spatio-Temporal Regularization", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* 2020.