



Article

A Fast Hyperspectral Tracking Method via Channel Selection

Yifan Zhang ¹, Xu Li ^{1,*}, Baoguo Wei ¹, Lixin Li ¹ and Shigang Yue ²¹ School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China² School of Computing and Mathematics Sciences, University of Leicester, Leicester LE1 7RU, UK; syue@lincoln.ac.uk

* Correspondence: lixu@nwpu.edu.cn

Abstract: With the rapid development of hyperspectral imaging technology, object tracking in hyperspectral video has become a research hotspot. Real-time object tracking for hyperspectral video is a great challenge. We propose a fast hyperspectral object tracking method via a channel selection strategy to improve the tracking speed significantly. First, we design a strategy of channel selection to select few candidate channels from many hyperspectral video channels, and then send the candidates to the subsequent background-aware correlation filter (BACF) tracking framework. In addition, we consider the importance of local and global spectral information in feature extraction, and further improve the BACF tracker to ensure high tracking accuracy. In the experiments carried out in this study, the proposed method was verified and the best performance was achieved on the publicly available hyperspectral dataset of the WHISPERS Hyperspectral Objecting Tracking Challenge. Our method was superior to state-of-the-art RGB-based and hyperspectral trackers, in terms of both the area under the curve (AUC) and DP@20pixels. The tracking speed of our method reached 21.9 FPS, which is much faster than that of the current most advanced hyperspectral trackers.

Keywords: channel selection; hyperspectral video; tracking speed



Citation: Zhang, Y.; Li, X.; Wei, B.; Li, L.; Yue, S. A Fast Hyperspectral Tracking Method via Channel Selection. *Remote Sens.* **2023**, *15*, 1557. <https://doi.org/10.3390/rs15061557>

Academic Editors: Fengchao Xiong, Jun Zhou, Yanfei Zhong, Pedram Ghamisi and Jocelyn Chanussot

Received: 5 February 2023

Revised: 8 March 2023

Accepted: 11 March 2023

Published: 12 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Object tracking has been a basic and active research topic in computer vision and pattern recognition, which has a wide variety of application fields and occupies an important position in both civilian and military applications [1–3]. The task of object tracking is to predict the motion state of a dynamic target in the subsequent frames according to the target's size and position in the initial or current frame of the video sequence. Object tracking methods can be roughly divided into generative methods and discriminant methods from the perspective of the observation model. Recently, discriminant tracking methods have gradually occupied the mainstream [4], which achieve satisfactory results with the help of correlation filtering [5–7] or deep learning [8]. Compared with the deep learning based object tracking methods, the correlation filtering based methods have great advantages in speed and they are widely used in applications requiring high real-time performance [9–11].

The traditional visual object tracker, also known as the RGB-based tracker, uses only three bands of visible light to track the target, which is still limited by some recognized factors, such as illumination change, detector jitter, background interference. It is not enough to describe the physical characteristics of the target in gray or color video, especially in the reflectivity of the material. RGB-based trackers often become vulnerable in complex scenes where the background is cluttered and the shape of the target changes significantly. The accurate description of target characteristics is the key to detecting and tracking. This problem can be effectively solved by object tracking in hyperspectral videos which provide joint spectral, spatial, and temporal information, enabling computer vision systems to perceive the materials of objects besides the shape, texture, and semantic relationship of objects. In recent years, with the maturity of hyperspectral imaging technology, hyperspectral

object tracking has gradually received widespread attention [12,13]. Hyperspectral imaging technology can provide dozens or even hundreds of spectral bands simultaneously for an interested scene. Hyperspectral data not only have the two-dimensional spatial features of the object, but also obtain very rich spectral information in the wavelength direction [14]. Hyperspectral videos can make some materials have outstanding manifestation in specific spectral channels. The deep characterization of target and background in spectrum is very helpful to improve the performance of tracking. Therefore, the reasonable extraction of the spectral features in hyperspectral object tracking can effectively improve the accuracy of tracking algorithm and reduce the target loss caused by weather conditions, equipment conditions, occlusion, camouflage, and other problems in complex environments [15].

With the gradual popularization of hyperspectral video, more scholars have participated in the research of hyperspectral video object tracking [16–18]. Hyperspectral object tracking technology has developed from tracking based on manual features to tracking based on deep features. In the early research, Banerjee et al. proposed the first hyperspectral video object tracker, in which they employed the spectral angle mapper to distinguish target and background [19]. Nguyer et al. combined the mean shift tracker and spectral reflectance to track the hyperspectral object [20]. However, when using the method, it is necessary to reduce the dimensionality of the spectral features, which requires a lot of calculation. In order to track multiple objects, Kandylakis et al. tried to estimate the background of the hyperspectral video sequences [21]. Recently, Qian et al. extracted the features of hyperspectral video by using the 3D patches from the target region as fixed convolution kernels (no training required) and proposed a convolutional network based hyperspectral tracking (CNHT) [22]. The method inputs the features into the kernelized correlation filter (KCF) tracker [23] rather than the histogram of oriented gradients (HOG) [24]. However, CNHT only considers fixed positive samples when training the filter and ignores the influence of negative samples on the tracker, resulting in poor tracking accuracy when background interference occurs in the task video. For aerial hyperspectral video, Uz Kent et al. designed a tracker based on deep hyperspectral KCF (DeepHKCF). The tracker uses VGG-Net deep neural network to learn both positive and negative samples, which can effectively improve the robustness of the tracker and significantly improve the tracking accuracy of aircraft in low frame rate video [25]. Since DeepHKCF separates hyperspectral video channels, the complete spatial structure information in hyperspectral video has not been fully explored, which limits the accuracy and robustness of the tracker in hyperspectral object tracking. Xiong et al. proposed the well-known material-based hyperspectral tracking (MHT) method which made full use of the spectral information of materials in hyperspectral video from three aspects: data set, material feature representation, and material-based tracking [26]. They combined the fractional abundances of the constituted material components and the spectral-spatial histogram of multidimensional gradients (SSHMG) as features, and then input them into the background-aware correlation filter (BACF) tracker [27]. The advantages of MHT make it a typical and competitive hyperspectral object tracking method. Li et al. introduced an attention-aware ensemble network (BAE-Net) for deep hyperspectral object tracking with the aid of a deep model trained on visual video for feature representation [28]. An anchor-free Siamese network (HA-Net) is proposed for hyperspectral object tracking by Liu et al., in which the spectral classification branch network from the anchor-free Siamese network is introduced [29]. This branch network can effectively identify objects and utilize the spectral information in the video sequences. However, the hyperspectral template online update module in HA-Net increases the computational burden obviously [30]. Furthermore, Li et al. proposed spectral-spatial-temporal attention neural network (SST-Net) [31]. SST-Net uses convolution and deconvolution structure as well as time attention with RNN structure to accurately describe the band relationship, thus effectively improving the tracking accuracy for hyperspectral object. Zhao et al. proposed a transformer-based fusion tracking network (TFTN), which is a hyperspectral and RGB fusion tracking framework [32]. TFTN constructs a dual-branch structure based on Siamese network to obtain modal specific representa-

tion of different modal images. Based on the general framework, the authors proposed Siamese 3-D convolutional neural network as a specific branch of hyperspectral mode. This is the first work to combine hyperspectral and RGB mode information to improve tracking performance, which achieved satisfactory tracking accuracy. So far, although the hyperspectral community has developed many tracking algorithms and started WHISPERS Hyperspectral Object Tracking Challenge in 2021 [33], most of the hyperspectral trackers are computationally expensive and the algorithms run slowly. How to increase the tracking speed is still a great challenge with practical implications. To deal with this problem, we propose a fast object tracking method based on the channel selection of hyperspectral video.

A preliminary version of this paper was published in WHISPERS2022 conference [34], in which the work was only a preliminary study of using channel selection to improve tracking speed. On the basis of the previous work, we carried out in-depth research in order to further boost the tracking speed while improving the tracking accuracy. Specifically, we use two-dimensional entropy instead of image entropy in the calculation of entropy module, and improve the BACF tracking framework to obtain a feature image with higher quality.

The main contributions of our work are as follows:

- (1) We design a channel selection strategy for the hyperspectral video and then input the selected channels into a BACF tracking framework, which successfully reduces the massive hyperspectral video input.
- (2) We combine the band-by-band HOG (BHOG) and SSHMG in the BACF and capture the local and global spectral features to obtain a feature image with higher quality, thus improving tracking accuracy.
- (3) Our method achieved the fastest tracking speed and the highest tracking accuracy on the only hyperspectral video object tracking benchmark dataset currently available [33].

The rest of this paper is organized as follows. In Section 2, we present the proposed method and describe the channel selection strategy and the improved BACF tracker in detail. In Section 3, we present the experiments and analyze the experimental results. Section 4 contains discussions. We make the conclusions in Section 5.

2. Proposed Method

The proposed method is shown in Figure 1, which consists of two parts—Part 1: Channel Selection, and Part 2: Improved BACF Tracker. Suppose the hyperspectral video has M channels and the n th frame image in each channel is denoted as $B_1^n, \dots, B_M^n$. Channel Selection inputs all the hyperspectral video channels and evaluates the quality of each one. Here the designed Channel Selection uses three special evaluation modules: contrast, entropy, and difference, and comprehensively selects the most valuable three channels as the input to the subsequent Improved BACF Tracker for object tracking. We describe the two parts in detail below.

2.1. Channel Selection

In contrast to traditional RGB video, hyperspectral video provides dozens or hundreds of video channels, bringing massive data to the calculation of object tracking. Due to the uniqueness of the spectral signature of the substance, the same target will show different appearances in different channels, which may lead to a different performance even for the same object tracker. On the other hand, it is necessary to remove the channels with redundancy related to special targets. In order to reduce computational burden and improve tracking speed, it is a feasible scheme to select the most representative few channels from hyperspectral video for object tracking. Therefore, evaluating the quality of each hyperspectral channel is a key issue. We focus on the intra-channel and inter-channel characteristics that are helpful for tracking, and design a simple but effective channel selection strategy. Unlike some existing hyperspectral video trackers that synthesize hyperspectral images into pseudo-color images as the input of RGB-based tracker, our channel selection strategy retains the hyperspectral semantic features of the visual

layers by evaluating the contrast, entropy, and target-background difference of the input hyperspectral channels. Specifically, the contrast and entropy modules are used to evaluate the spatial characteristics of individual channel image, while the difference module focuses on the spatial-spectral characteristics between channels. Finally, the candidate channels are obtained by comprehensive use of these three modules.

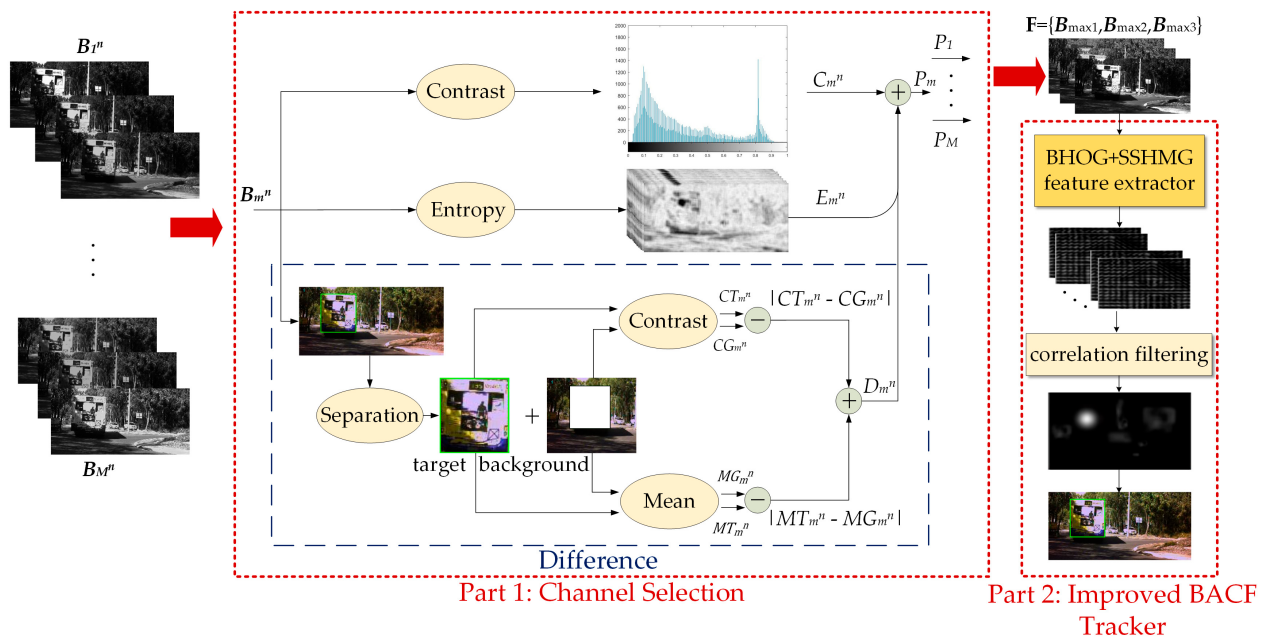


Figure 1. The architecture of the proposed method. The method mainly consists of two parts—Part 1: Channel Selection, and Part 2: Improved BACF Tracker. Part 1 is mainly composed of contrast module, entropy module, and difference module. Channel Selection is responsible for selection of the most representative few candidate channels from the input hyperspectral video. Part 2 is used to track the object in the images of the selected channels F . In Part 2, BHOG + SSHMG is designed to improve the classical BACF tracker.

2.1.1. Contrast Module

The salient information in an image can be described by image contrast. Hence, we design a new contrast module by calculating the local contrast of an image to perceive the saliency information. Our contrast module can reflect some spatial structure features in the channel image. We denote the contrast module as C_m^n which is defined as:

$$C_m^n = \frac{\sum_{i \in B_m^n} \sum_{j \in w_i} (B_{m,j}^n - B_{m,i}^n)^2}{S} \quad (1)$$

where $B_{m,i}^n$ represents the pixel value of the pixel position i in the channel image B_m^n , and w_i denotes the four-neighborhood (top, bottom, left, and right) centered at i . j is the pixel in w_i . S is the total number of the operations of the numerator part. The larger the value of C_m^n , the greater the local gray difference, indicating that the channel image B_m^n is clearer.

2.1.2. Entropy Module

We believe that using the channels with high information entropy will be more beneficial to object tracking. The two-dimensional entropy of the channel image can reflect the spatial characteristics of the image gray distribution. The richer the entropy information in

an image, the more valuable the extracted features are. We denote the two-dimensional entropy of the n th frame hyperspectral image in the m th channel as E_m^n .

$$E_m^n = - \sum_{i \in B_m^n} \sum_{j \in w_i} P_{i,j} \lg(P_{i,j})$$

$$P_{i,j} = \frac{f(B_{m,i}^n, B_{m,j4}^n)}{wh} \quad (2)$$

where $B_{m,i}^n$ represents the pixel value at position i , $B_{m,j4}^n$ represents the average value of pixels in the four neighboring regions of position i . $f(i, j)$ is the number of occurrences of feature binary $(B_{m,i}^n, B_{m,j4}^n)$ in the image B_m^n , and w and h are the width and height of the image, respectively.

2.1.3. Difference Module

Highlighting the difference between the target and the background in a hyperspectral image helps to represent and retain the semantic features between the video channels. Therefore, we design a difference module to evaluate the difference between the target and the background in the channel image. The more obvious the difference, the higher the target detection accuracy in tracking. We use mean and contrast to measure the spatial and spectral changes of the target area and the background area.

First, separate the target area and the background area from a channel image according to the ground truth. Then calculate the mean and the contrast of the two areas. The contrast calculation follows Equation (1). The final difference D_m^n of the n th frame in the m th channel is defined as:

$$D_m^n = |MT_m^n - MG_m^n| + |CT_m^n - CG_m^n| \quad (3)$$

in which MT_m^n and MG_m^n represent the mean value of the target area and the background area, respectively. CT_m^n and CG_m^n are the contrast of the two areas, respectively.

2.1.4. The Candidate Channels Selection

Existing hyperspectral video target-tracking tasks usually use fixed-angle cameras to shoot and then track the target, so the similarity of the background environment in a single task video will be high. If only the first frame of the video is entered in Channel Selection, the noise interference cannot be ignored. However, if too many video frames input, the amount of the computation will increase significantly. By evaluating the quality of a few initial frames, the representative channels can be selected out. Here we use the first five frames of the task video to select channels, which can effectively eliminate interference without affecting the tracking speed.

We comprehensively use the modules of contrast, entropy, and difference, and build an overall evaluation index (denoted as P_m) to select the most valuable channels. The index P_m of the first n frames of the channel B_m is defined as follows:

$$P_m = \sum_{n=1}^5 (\alpha C_m^n + \beta E_m^n + \gamma D_m^n) \quad (4)$$

where α , β , and γ are the weight coefficients of the different modules, respectively. The value of n ranges from 1 to 5, which means Channel Selection only considers the first five frames of the hyperspectral video input.

Hyperspectral video has a large number of channels and contains rich spatial-spectral information. However, useful spectral information related to specific targets is limited. We believe that a few channels can represent the most useful information. As the first part of our method, Channel Selection selects the three most valuable channels as the input of the subsequent BACF Tracker for object tracking. The first three channels with the largest P_m values are selected as the candidates, denoted as F .

$$F = \{B_{max1}, B_{max2}, B_{max3}\} \quad (5)$$

where $B_{max1}, B_{max2}, B_{max3}$ are the three channels selected from the M hyperspectral channels $\{B_1, B_2, \dots, B_M\}$.

2.2. BACF Tracker

2.2.1. Classical BACF Tracker

Discriminant correlation filter is widely used in object tracking because of its fast computational speed realized by fast Fourier transform. As an object tracker based on correlation filtering, BACF has high accuracy and excellent speed. BACF treats all background patches as negative samples by using a rectangular mask covering the central part of circular samples [27], thus obtaining more real samples from the background.

The classical BACF learns the filter h_k by the following objective function:

$$E(h) = \frac{1}{2} \sum_{t=1}^T \left\| y(t) - \sum_{k=1}^K h_k^\top P x_k [\Delta \tau_t] \right\|_2^2 + \frac{\lambda}{2} \sum_{k=1}^K \| h_k \|_2^2 \quad (6)$$

where x_k and h_k refer to the feature image and the filter of the k th channel, respectively. K represents the number of channels of the feature image, and T is the feature dimension. y is the desired correlation response and $y(t)$ is the t th element of y . λ is a regularization operator. P is the binary matrix for clipping the central patch of x_k . $[\Delta \tau_t]$ is a cyclic shift operator. $x_k[\Delta \tau_t]$ indicates that t step discrete cyclic shift is applied to the feature x_k .

2.2.2. Improved BACF tracker

It is commonly known that the quality of the feature image plays a decisive role in the accuracy of object tracking. For the feature image x , the classical BACF is obtained by histograms of oriented gradients (HOG) feature extraction. However, HOG feature only considers the spatial information and ignores the rich spectral information in hyperspectral images, resulting in poor tracking results for hyperspectral video. Fortunately, there are some new variants based on HOG features for hyperspectral images. In DeepHKCF method, a band-by-band HOG is designed which adopts HOG feature extraction in each hyperspectral band and considers the global spectral features of all bands [25]. To highlight the spectral information of materials, Xiong et al. took into account the local spectral features of the target and developed a spectral-spatial histogram of multidimensional gradients (SSHMG) to improve the quality of feature images [26]. Inspired by this, we combine BHOG and SSHMG to extract both local and global spectral features from hyperspectral channel images. Specifically, for the channel images, we first use the BHOG operator to obtain the feature image considering the global spectrum, and then send the feature image to SSHMG to produce the final feature image characterized by local spectral information. Thus, BHOG + SSHMG extracts both the local and global spectral features from the hyperspectral images to cope with complex scenes, such as fast moving targets, camouflaged targets, background clutter, and so on. The improved BACF tracker uses BHOG + SSHMG as the feature extractor to improve the quality of the feature image, which help to improve the accuracy of object tracking. The corresponding experiments further demonstrate the effectiveness of our improved BACF tracker. Section 3 gives the experimental results in detail.

Using the characteristics that cyclic samples can be quickly solved in the frequency domain, Equation (5) can be converted into frequency domain to process the feature image extracted by BHOG + SSHMG, and the formula is as follows:

$$E(h) = \frac{1}{2} \| \hat{y} - \hat{X} \sqrt{T} (F F^\top \otimes I_K) h \|_2^2 + \frac{\lambda}{2} \| h \|_2^2 \quad (7)$$

where X is the $T \times KT$ matrix, defined as $X = [\text{diag}(x_1), \dots, \text{diag}(x_K)]$. F is orthogonal $T \times T$ matrix, which maps any T -dimensional vectorized signal to Fourier domain. I_K is $K \times K$ identity matrix. $\hat{\cdot}$ and $\otimes$ represent discrete Fourier transformation and the Kronecker product, respectively.

The auxiliary matrix $\hat{\mathbf{g}} = \sqrt{T}(\mathbf{F}\mathbf{P}^\top \otimes \mathbf{I}_K)\mathbf{h}$ is introduced here, and the Lagrangian augmented function is shown as:

$$\mathcal{L}(\hat{\mathbf{g}}, \mathbf{h}, \hat{\boldsymbol{\zeta}}) = \frac{1}{2} \|\hat{\mathbf{y}} - \hat{\mathbf{X}}\hat{\mathbf{g}}\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}\|_2^2 + \hat{\boldsymbol{\zeta}}^\top (\hat{\mathbf{g}} - \sqrt{T}(\mathbf{F}\mathbf{P}^\top \otimes \mathbf{I}_K)\mathbf{h}) + \frac{\mu}{2} \|\hat{\mathbf{g}} - \sqrt{T}(\mathbf{F}\mathbf{P}^\top \otimes \mathbf{I}_K)\mathbf{h}\|_2^2 \quad (8)$$

where μ is the penalty factor and $\hat{\boldsymbol{\zeta}}$ is the $\text{KT} \times 1$ Lagrange vector in Fourier domain. Equation (7) can be solved iteratively using Alternating Direction Method of Multipliers (ADMM) [35] technique. For each subproblem, $\hat{\mathbf{g}}^*$ and $\mathbf{h}^*$ have the closed form solutions. The optimization subproblem $\hat{\mathbf{g}}^*$ and $\mathbf{h}^*$ is omitted here. Readers are advised to read [27] for more details about the optimization process.

The update strategy of the BACF model is the same linear interpolation method as the traditional correlation filter:

$$\hat{\mathbf{x}}_{\text{model}}^{(f)} = (1 - \eta)\hat{\mathbf{x}}_{\text{model}}^{(f-1)} + \eta\hat{\mathbf{x}}^{(f)} \quad (9)$$

where η is the adjustment ratio, and f is the current frame.

3. Experiment and Results

In this section, we first describe the hyperspectral video dataset and experiment setting, and then show the ablation studies on different channel selection strategies and feature extractors, respectively. Finally, we compare our method with both RGB-based trackers and hyperspectral object trackers. The parameters of the compared methods are all automatically set the same as the recommendations in their original literature.

3.1. Dataset

As shown in Figure 2, the dataset used in this study is the publicly available hyperspectral dataset released as the part of the WHISPERS Hyperspectral Object Tracking Challenge [33]. The hyperspectral video sequence, false-color video sequence and RGB video sequence constitute the competition dataset. Each video type contains 50 video sequences, with an average of 425 frames per video sequence. Hyperspectral videos were acquired by using XIMEA Snapshot VIS camera with 16 channels at 25 FPS. Each frame is initially captured in 2D format, with 16 channels arranged in a mosaic pattern. The hyperspectral bands cover the range from 470 nm to 620 nm and each band image originally consists of 512×512 pixels. The ground truth values of targets in all videos of the dataset are manually marked by Xiong et al. [26].

3.2. Experiment Setting

In all experiments, the parameter of learning rate is set to 0.004. The weight coefficients in Channel Selection are set as $\alpha = 2$, $\beta = 1$, $\gamma = 1$. All other parameters are set the same as those in BACF [27]. We conducted all experiments under Matlab2020 in the computing environment of Intel (R) Core (TM) i9-10900K CPU and NVIDIA GeForce RTX 3060 GPU. We use the accuracy plots, success plots, AUC score, and DP score to evaluate the tracking performance of different trackers.

A precision plot records the fraction of the frames whose estimated locations is within a given threshold distance to the ground-truth centers. The average distance precision (DP) rate is reported at a threshold of 20 pixels (DP@20pixels). The location error calculation formula is as follows:

$$\text{Loc}(G, O) = \left(\sum_{d=1}^D |G_d - O_d|^2 \right)^{1/2} \quad (10)$$

where D represents the dimension information of the target, which is generally set to 2 in the image field, G_d is the center position coordinate of the real target, and O_d is the center position coordinate of the target predicted by the object tracker. The location error threshold

is the boundary between the real position of the target and the position error predicted by the tracker. When the location error of the current frame is less than the location error threshold, the tracking of the frame is considered accurate.

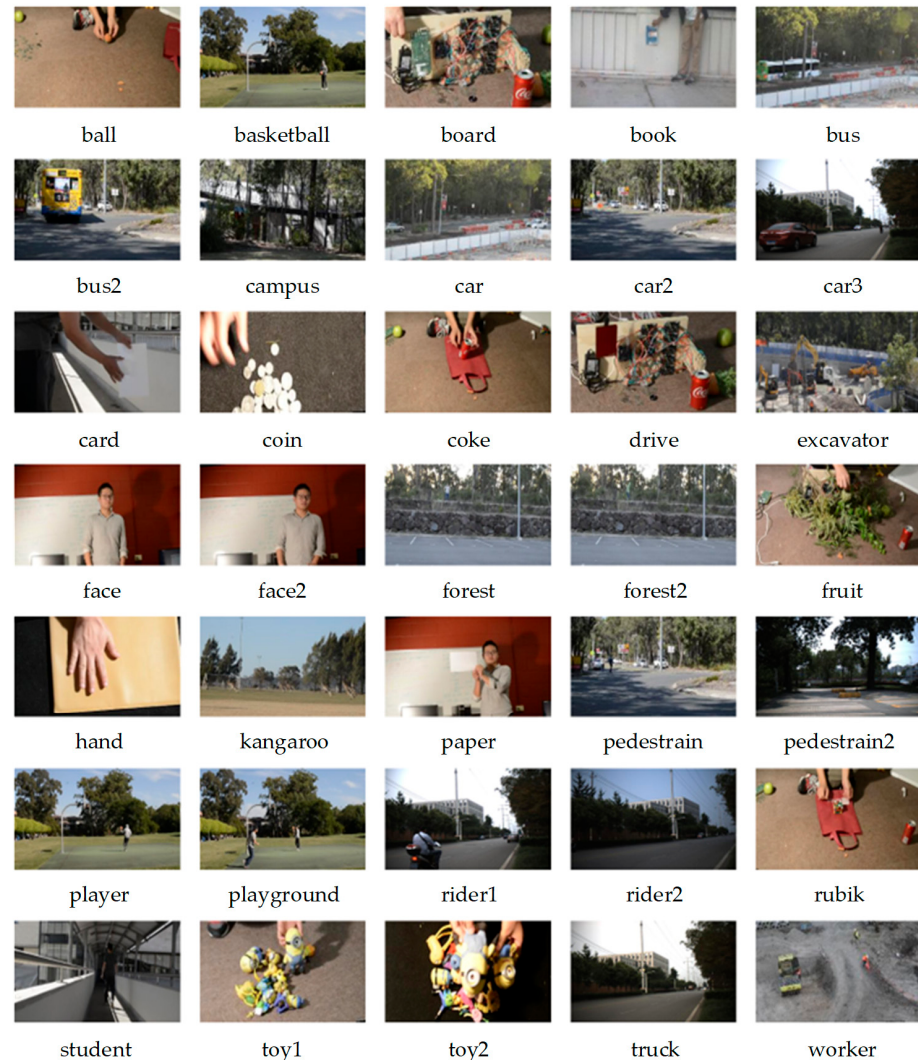


Figure 2. Hyperspectral video dataset scenes [33].

A success plot describes the percentage of successful frames whose overlap ratio between the predicted bounding box and the ground truth is larger than a certain threshold, which varies from 0 to 1. The overlap rate is defined as:

$$OS = \frac{|B_{pre} \cap B_{gt}|}{|B_{pre} \cup B_{gt}|} \quad (11)$$

where B_{pre} is the target prediction box, and B_{gt} is the real position of the target. The overlap threshold is the boundary of the overlap ratio. When the overlap ratio is larger than the overlap threshold, this frame is considered to be successful. The overlap threshold value range is 0–1. The larger the threshold value, the lower the success rate. A success rate graph curve can be obtained by referring to different threshold values. AUC score is the area of the lower curve of the success rate curve.

All the results are presented with one-pass evaluation (OPE); that is, the tracker runs throughout the test sequence and is initialized from the ground-truth location in the initial frame. The evaluation indexes employed are the same as those suggested in [36] used for the object tracking benchmark.

3.3. Comparison of Different Channel Selection Strategies

As shown in Figure 1, three modules, i.e., contrast, entropy, and difference, make up of Channel Selection and they work together to select the candidate channels from the input hyperspectral video. We use the ablation experiments to demonstrate the effectiveness of our channel selection strategy. The contrast-based strategy, entropy-based strategy, difference-based strategy, and their combinations are tested individually, and then all results are compared comprehensively. It needs to be noted here that all the subsequent tracking steps are the same except for the difference strategy tests.

Figure 3 reports the object tracking performances using different channel selection strategies. Figure 3a,b are the precision plot and success plot of channel selection strategies, respectively. We narrow the ordinate range of the experimental result plots to make the accuracy difference of each channel selection strategy more obvious. On the whole, the trend of performances of all strategies is consistent. However, the curves display obvious differences in the 10–50 interval of abscissa in the precision plots and the 0–0.6 interval of abscissa in the success plots. The reason is that different channel selection strategies lead to different rationalities of the selected channels. In the plots, the higher the curve position, the better the performance of the corresponding strategy. The contrast-based channel selection strategy displays the worst accuracy among all the tested strategies. The reason is the contrast-based strategy only takes into account the brightness of the image, but ignores the information richness and the inter-channel characteristics. Although the performance the entropy-based strategy is higher than that of the contrast-based, its accuracy is not as good as the other strategies due to its one-sided consideration. The performance of difference-based strategy is better than that of the contrast- or entropy-based, which indicates the inter-channel features play an important role in hyperspectral tracking. Adding contrast index on the basis of target background difference index can effectively improve the accuracy of object tracking. However, when the entropy indicator is combined with other indicators, the tracking accuracy does not improve, since it is not compatible with the contrast or difference indicator alone. In all tests, the strategy of contrast + entropy + difference achieves the highest accuracy, which proves the advantages of our channel selection strategy.

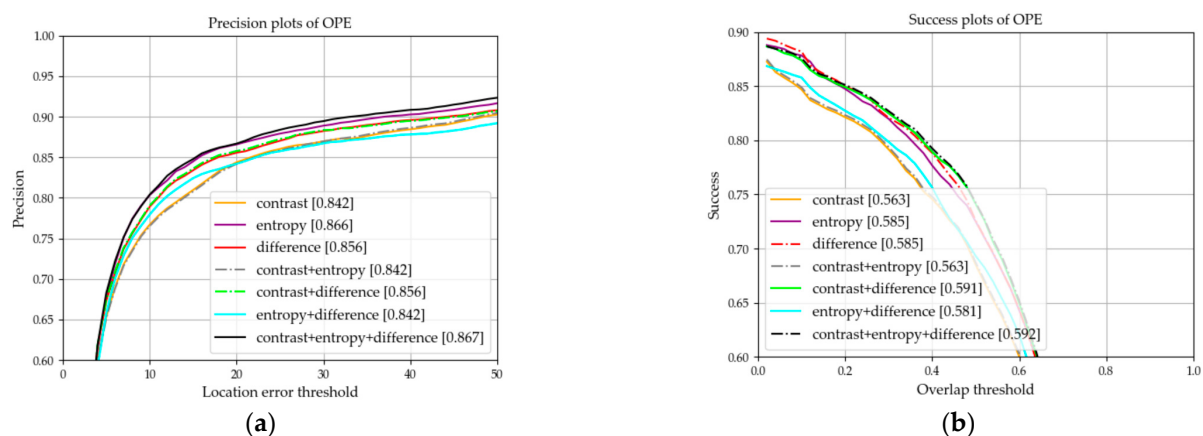


Figure 3. Performances of different channel selection strategies. (a) Precision plots of different channel selection strategies; (b) Success plots of different channel selection strategies.

Table 1 lists AUC and DP@20pixels values of different channel selection strategies. From the results of the AUC index, difference-based strategy has a greater impact on performance than contrast strategy and entropy strategy. In terms of AUC, the channel selection strategy of contrast + entropy + difference achieves the highest detection accuracy of 0.592, which shows that this strategy selects the channels most conducive to object tracking. In terms of DP@20pixels, the strategy of contrast + entropy + difference reaches 0.867, which verifies correctness of Channel Selection combining three modules. The

strategy of contrast + entropy + difference achieves the best according to both AUC and DP@20pixels, which demonstrates the rationality and effectiveness of the proposed channel selection strategy.

Table 1. Tracking performance comparison of channel selection strategies. The best two results are marked with red and blue.

Strategy	AUC	DP@20pixels
contrast	0.563	0.842
entropy	0.585	0.866
difference	0.585	0.856
contrast + entropy	0.563	0.842
contrast + difference	0.591	0.856
entropy + difference	0.581	0.842
contrast + entropy + difference	0.592	0.867

3.4. Comparison of Feature Extractors

Feature extraction is a key point in the object tracking process. The better the quality of the extracted feature image, the higher the tracking accuracy. The classical BACF tracker uses HOG as the feature extractor which only considers the local spatial features in visual images. For hyperspectral object tracking, BHOG and SSHMG feature extractors are the development and evolution products of BACF, which take into account the global and/or local spectral information in hyperspectral images. We combine BHOG and SSHMG to obtain the feature image with higher quality. In order to verify the effectiveness, it is necessary to conduct ablation experiments on different feature extractors. In this experiment, we test the performance of seven feature extractors, including HOG, BHOG [25], SSHMG [26], HOG combined with BHOG (HOG + BHOG), HOG combined with SSHMG (HOG + SSHMG), BHOG combined with SSHMG (BHOG + SSHMG), and HOG + BHOG + SSHMG. In all tests, except for the different feature extraction methods in BACF, all other tracking steps strictly follow the original BACF framework.

Figure 4 shows the object tracking performance on different feature extractors. Figure 4a,b are the precision plots and success plots of feature extractors, respectively. We narrow the ordinate range of the experimental result plots to make the accuracy difference of each feature extraction method more obvious. On the whole, the trend of all curves is similar. There is a distinction shown between curves in the 10–50 interval of abscissa in the precision plots and the 0–0.6 interval of abscissa in the success plots. This is because different feature extractors have different feature extraction capabilities. HOG feature extractor provides the worst accuracy in all extractors, since the original spectrum is very sensitive to light changes. BHOG considers the global spectral information of all channels to provide better performance. SSHMG makes use of the local spectral-spatial structure information, so it outperforms HOG and BHOG. It is easy to find that the object tracking accuracy can be improved by combining different single feature extractors. The combinations are effective and feasible for high accuracy. Among them, our BHOG + SSHMG combination achieves the best results.

Table 2 reports the performances of different feature extractors. In terms of AUC, the feature extraction method of BHOG + SSHMG achieves the highest accuracy of 0.608, which shows that combining global spectral information with local spectral information for feature extraction in hyperspectral video has advantages. With reference to DP@20pixels, SSHMG achieves the highest accuracy of 0.904, and BHOG + SSHMG achieves 0.903. This shows that using local spectral information has great advantages in average distance accuracy. For individual feature extractor, HOG runs the fastest, while SSHMG is the slowest. For the combination of feature extractors, the more combinations, the slower the operation speed. HOG + BHOG + SSHMG runs the slowest. It is worth noting that the BHOG + SSHMG feature extractor achieves the highest accuracy, which proves that

the method combining local spectral information extraction and global spectral feature extraction is more conducive to object tracking in hyperspectral video.

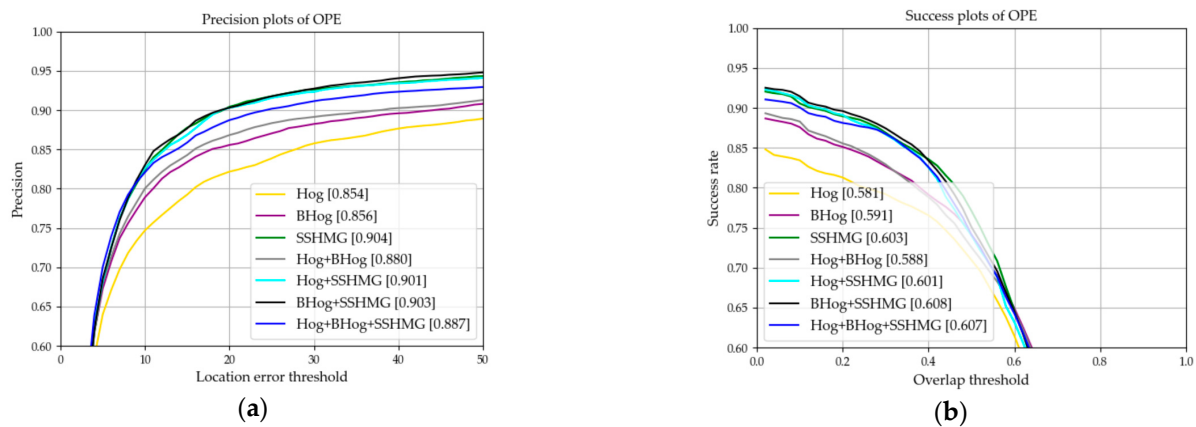


Figure 4. Performance of different feature extractors. (a) Precision plots of different feature extractors; (b) Success plots of different feature extractors.

Table 2. Tracking performance comparison of feature extractors. The best two results are marked with red and blue.

Feature Extractor	AUC	DP@20pixels	FPS
HOG	0.581	0.854	86.460
BHOG [25]	0.591	0.856	48.683
SSHMg [26]	0.603	0.904	39.106
HOG + BHOG	0.588	0.880	33.829
HOG + SSHMG	0.601	0.901	32.671
BHOG + SSHMG	0.608	0.903	21.928
HOG + BHOG + SSHMG	0.607	0.887	20.576

In order to display the effectiveness of different feature extractors more intuitively, Figure 5 shows the response results of the classical BACF and our improved version. Figure 5b,c depict the responses of HOG feature image and BHOG + SSHMG feature image, respectively. When the color difference between the background region and the target region is larger, the tracking results have higher robustness and accuracy. It can be seen that when using the HOG feature, the color difference between the target region and the background region is small. The yellow dots occupy a large proportion in the red background region, which indicates that the HOG feature has low accuracy. In contrast, the color difference of BHOG + SSHMG is very large, in which the yellow dots are fewer and concentrated around the target region. Therefore, our improved BACF has higher accuracy and robustness than the classical.

3.5. With and Without Channel Selection Strategy

In this experiment, we explore the impact of our channel selection strategy on hyperspectral object tracking. For the same object tracker, we compare the tracking results of using our channel selection strategy and not. The tracking method is based on our improved BACF in which BHOG + SSHMG is used as the feature extractor.

Figure 6 shows the tracking performance comparison. As is shown in the figure, tracking with our strategy is the black curve, and the yellow curve is the tracking without any channel selection. It can be seen that the black curve is higher than the yellow one, which means tracking with our channel selection strategy has obvious advantages in both success rate and the accuracy. Tracking without channel selection considers too much spectral information in feature extraction, which leads to information redundancy and reduces the accuracy.

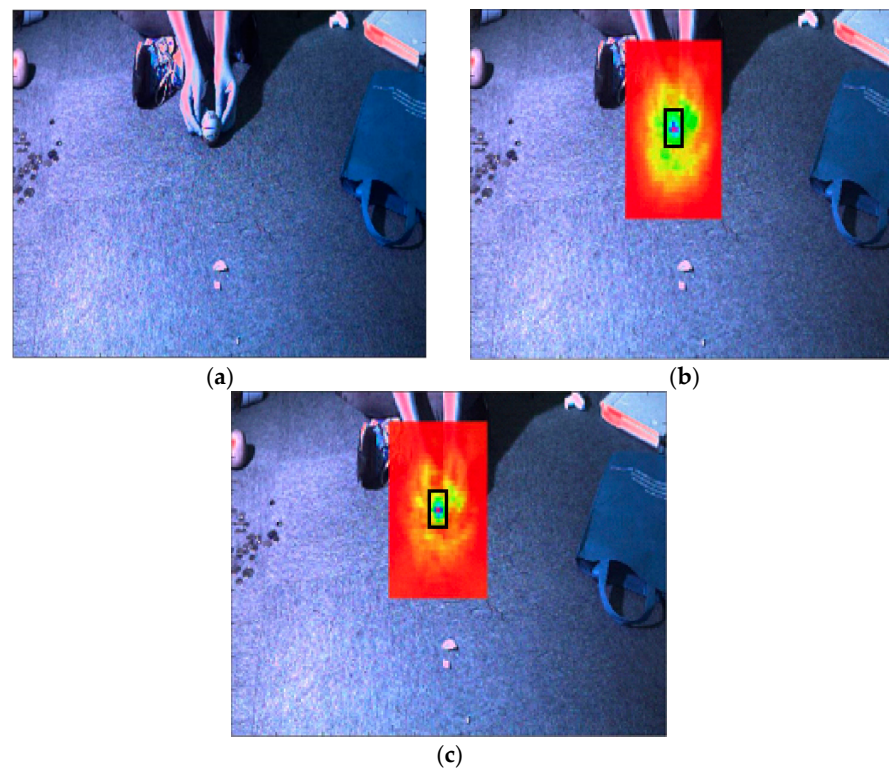


Figure 5. Response of different feature extractors. (a) Original image; (b) Response of HOG; (c) Response of BHOG + SSHMG. In the figure, the red rectangular area is the search range of the tracker for the target at the current time, i.e., the background area. The black rectangle represents the location of the target predicted by the tracker, which is the target area. Within the predicted box, the closer the pixel is to green, the more likely it is to belong to the target. When the color difference between the background area and the target area is large, it indicates that the tracker is robust in tracking the targets.

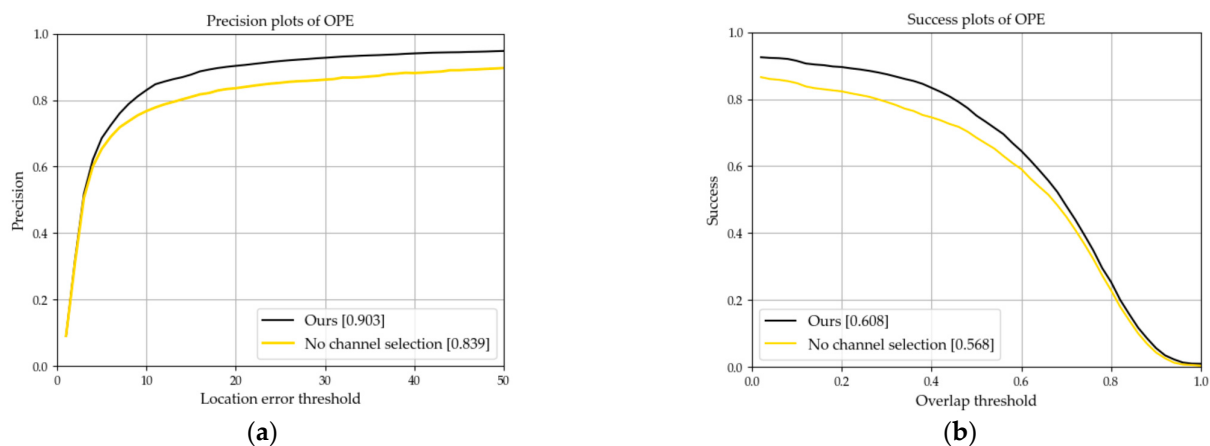


Figure 6. Comparison of with and without channel selection strategy. (a) Precision plots of trackers with/without channel selection strategy; (b) Success plots of trackers with/without channel selection strategy.

Table 3 reports the objective scores of the tracking results. Our channel selection helps AUC and DP@20pixel increase by nearly 7.04% and 7.63%, respectively. More importantly, the tracking speed has been greatly improved. FPS increased from 6.376 to 21.928, achieving an increase of 244%. From the comparison experiments, we can conclude that the channel selection is an effective way to improve the tracking speed for hyperspectral

video. Furthermore, our channel selection strategy is effective in both tracking accuracy and speed improvement.

Table 3. Tracking performance comparison with and no channel selection.

Trackers	AUC	DP@20pixels	FPS
Ours	0.608	0.903	21.928
No channel selection	0.568	0.839	6.376

3.6. Quantitative Comparison with RGB Object Trackers

With the help of the rich spectral information in hyperspectral videos, hyperspectral object tracking can effectively address the problems encountered in RGB object trackers and improve the tracking accuracy. Therefore, we compare the proposed tracking method with some existing RGB object trackers to verify its ability to face complex situations. Since all RGB trackers are designed for traditional color video, we use false color generated by hyperspectral video to test them in the comparison experiment.

In this experiment, we compare our tracker with five advanced RGB-based trackers, including BACF [27], KCF [23], DSST [37], C-COT [38], and CF-Net [39]. KCF, DSST, and BACF are object trackers based on manual feature extraction, while C-COT and CF-Net are based on depth feature extraction. KCF is a kernel correlation filter in which kernel techniques are applied to realize nonlinear classification boundary. DSST describes a scale adaptive tracking method, in which two discriminant correlation filters are learned for target position and scale estimation, respectively. BACF aims to reduce the KCF boundary effect caused by the periodic assumption on training samples by regularizing the correlation filter according to the spatial distribution. Deep learning-based object trackers represent another kind of advanced tracking methods. C-COT trains the convolution operator in the continuous space domain to achieve the integration of multi-resolution feature maps. CF-Net is an end-to-end tracker in which the correlation filter is interpreted as a differentiable layer in a deep neural network.

Table 4 reports the tracking performance of all compared methods. The results show that the KCF method gives unsatisfactory AUC and DP@20pixels scores due to the limited consideration of the size estimation. When an object has the appearance similar to the background, it cannot detect keys. BACF [27] integrates background information to learn more discriminant filters, so as to obtain better tracking performance. Our method and C-COT achieve the best two AUC scores of 0.608 and 0.557, respectively. Compared with BACF, the AUC and DP@20pixels scores of our method is improved by 6.4% (0.544 to 0.608) and 8.7% (0.816 to 0.903), respectively. It demonstrates that our improved BACF is effective. In addition, when compared with other trackers, our tracker ranks first in a series of thresholds, which means that our method can better utilize the rich spectral information in hyperspectral video. The reason for RGB object trackers obtaining lower performance is that the tested hyperspectral dataset contains scenes with camouflaged targets or scenes with background clutter. The improvement of tracking performance requires more spectral information, which is ignored by the RGB-based trackers.

Table 4. Tracking performance comparison with RGB object trackers. The best two results are marked with red and blue.

Trackers	AUC	Δ AUC	DP@20pixels	Δ DP
Ours	0.608	+6.4%	0.903	+8.7%
BACF [27]	0.544	-	0.816	-
KCF [23]	0.408	−13.6%	0.583	−23.3%
DSST [37]	0.442	−10.2%	0.705	−11.1%
C-COT [38]	0.557	+1.3%	0.869	+5.3%
CF-Net [39]	0.543	−0.1%	0.872	+5.6%

3.7. Hyperspectral Trackers Comparison

In this experiment, we compare with CNHT [22], DeepHKCF [25] and MHT [26] hyperspectral trackers to verify the performance of our method. Figure 7 depicts the evaluation results of all hyperspectral trackers in precision plot and success plot. Figure 7a,b are the precision plots and success plots of hyperspectral trackers, respectively. The higher the curve position, the better the performance of the tracker. It can be seen that the performance of CNHT is worst and our tracker achieves the best. Since DeepHKCF does not fully consider the spectral characteristics of hyperspectral video, it results in low accuracy. Although the performance of MHT method is close to that of ours, there is still a gap between them.

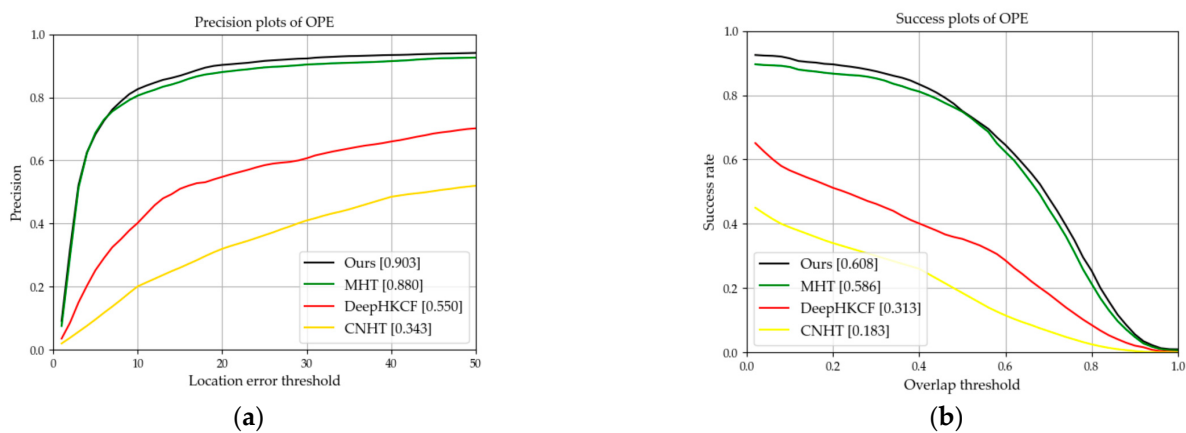


Figure 7. Comparison of hyperspectral trackers. (a) Precision plots of different hyperspectral trackers; (b) Success plots of different hyperspectral trackers.

Table 5 lists the numerical evaluation results of all hyperspectral trackers. CNHT obtains the worst AUC score of 0.183, because it only considers the fixed positive samples when learning the convolution filter. Compared with CNHT, DeepHKCF uses positive and negative samples to learn discriminant feature representation. Therefore, the AUC and DP@20pixels scores of DeepHKCF increase to 0.313 and 0.550, respectively. Among all trackers, our tracker has the best accuracy performance. Compared with the MHT method, the AUC and DP@20pixels values of our method are improved by 2.2% (0.586 to 0.608) and 2.3% (0.880 to 0.903), respectively. This proves that our tracker is suitable and effective for object tracking in hyperspectral video. In terms of running speed, MHT, DeepHKCF, and CNHT have similar FPS, belonging to the same order of magnitude. The running speed of our method has changed significantly, reaching 21.928 FPS, more than twice as fast as MHT. The reason for their low running speed is that both DeepHKCF and MHT use convolutional networks and CNHT has too many convolution layers when learning the convolution filter. In contrast, our proposed channel selection strategy can greatly reduce the amount of data used for tracking calculation and significantly improve the running speed.

Table 5. Tracking performance comparison with hyperspectral trackers. The best two results are marked with red and blue.

Trackers	AUC	Δ AUC	DP@20pixels	Δ DP	FPS	Δ FPS
Ours	0.608	+2.2%	0.903	+2.3%	21.928	+13.343
MHT [26]	0.586	-	0.880	-	8.585	-
DeepHKCF [25]	0.313	−27.3%	0.550	−33.0%	7.965	−0.620
CNHT [22]	0.183	−40.3%	0.343	−53.7%	8.101	−0.484

3.8. Visual Comparison with Hyperspectral Trackers

We select six video sequences (ball, basketball, coke, paper, fruit, and kangaroo) to compare four hyperspectral trackers including CNHT, DeepHKCF, MHT, and our method. These video sequences have different typical difficulties, such as occlusion, fast motion, rotation, and so on. Figures 8–13 show the visual tracking results of the hyperspectral trackers.

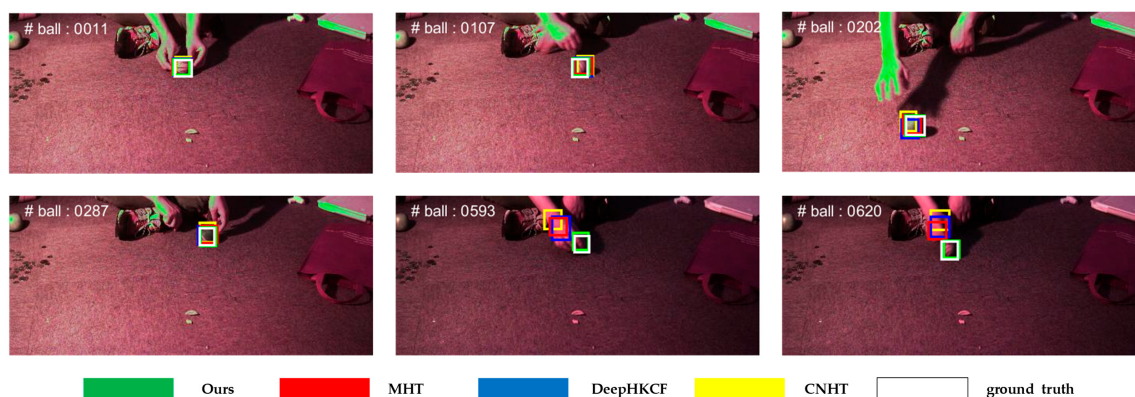


Figure 8. Tracking result of the ball sequence.

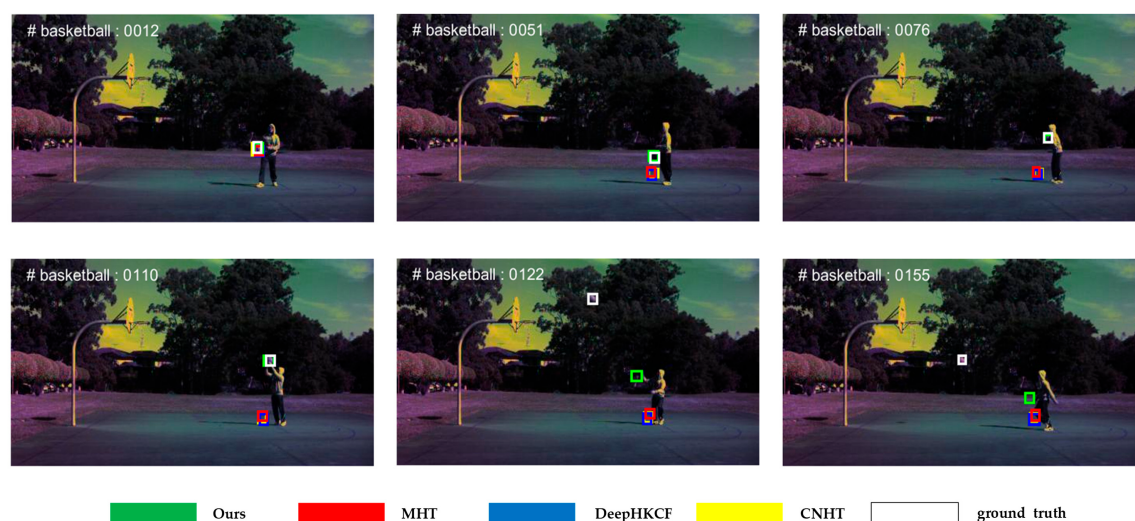


Figure 9. Tracking result of the basketball sequence.

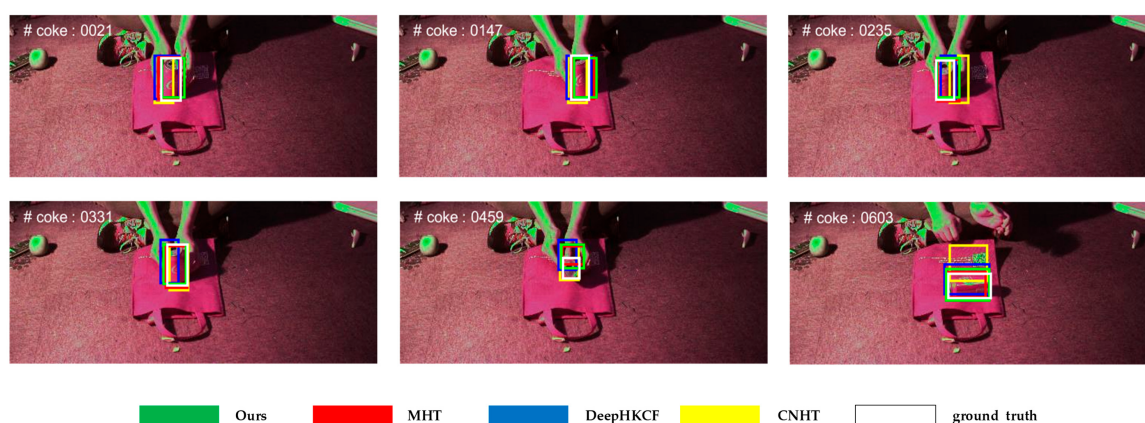


Figure 10. Tracking result of the coke sequence.

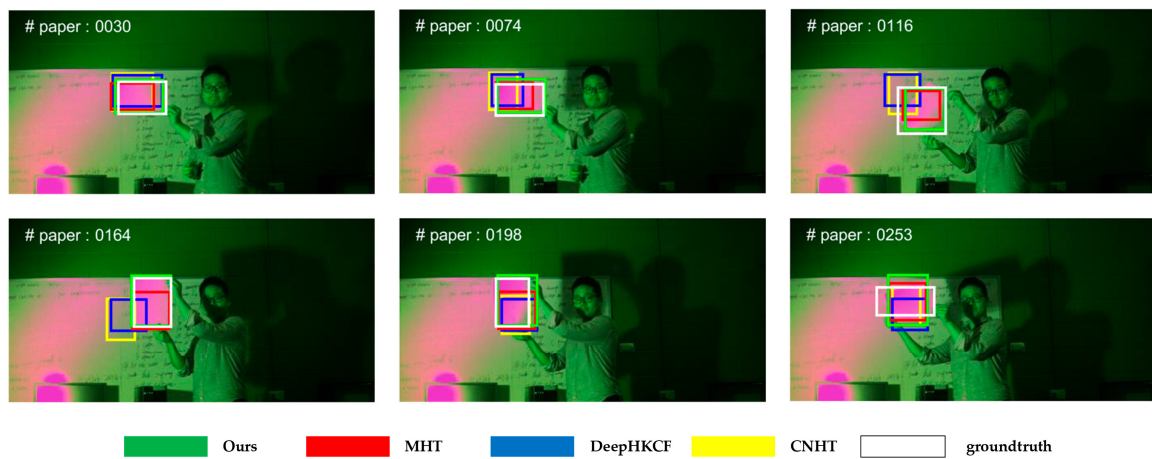


Figure 11. Tracking result of the paper sequence.

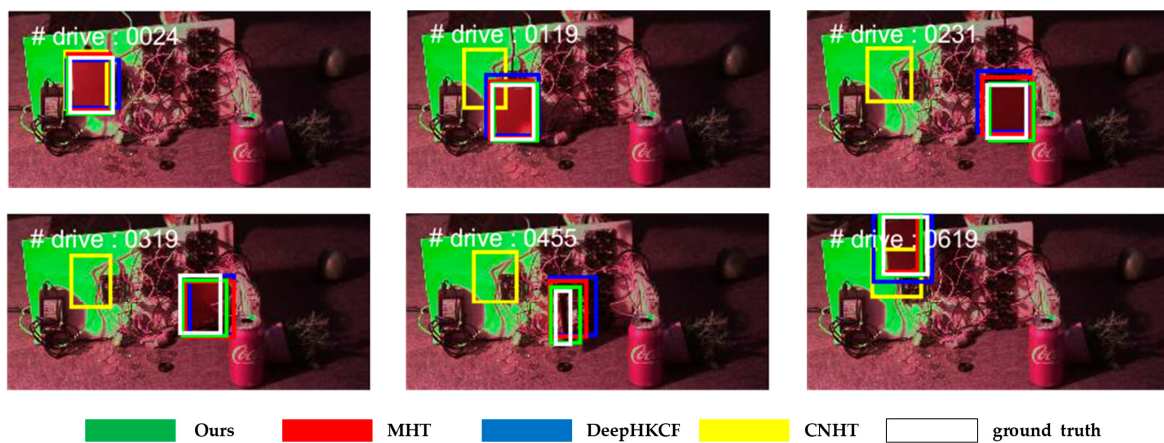


Figure 12. Tracking result of the drive sequence.

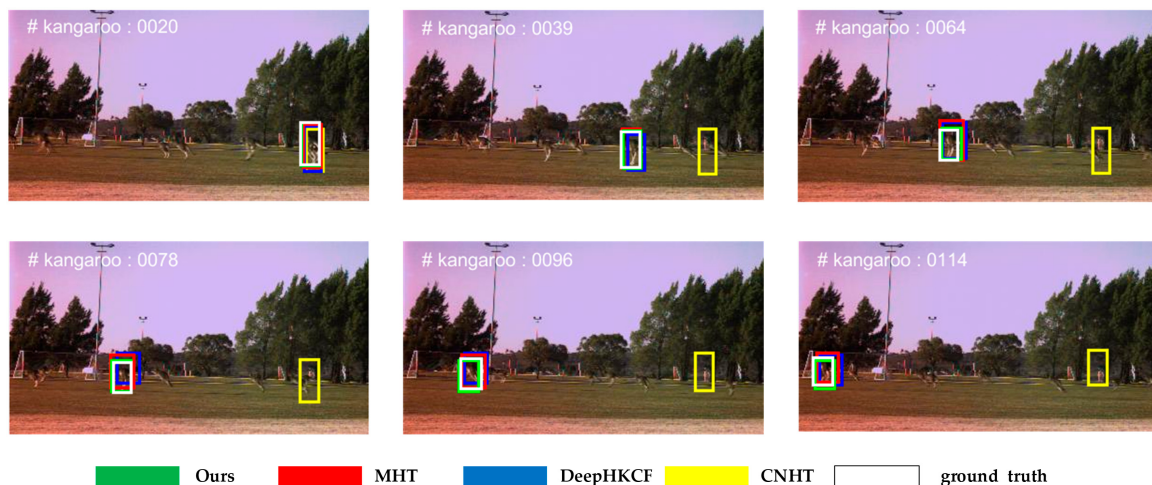


Figure 13. Tracking result of the kangaroo sequence.

Figure 8 shows the tracking results of the ball video sequence. Occlusion is the difficulty of this video. From the tracking results, it can be seen that occlusion happens in the middle sequences. In the initial tracking, all trackers can accurately track the target. However, after the hand completely covers the ball in the video, all the compared trackers mistakenly use the hand as the feature to update the template, resulting in the tracking of

the hand as the target in the subsequent video, which ultimately leads to the reduction of tracking accuracy. Our tracker correctly updates the template after the complete occlusion, and successfully continues to track the ball after the ball reappears, thus achieving higher tracking accuracy.

Figure 9 shows the tracking results of the basketball video sequence. The basketball sequence shows the tracking of a basketball in a daily scene whose difficulties of tracking are the fast moving small targets and the background interference. From the tracking results, we can see that the compared trackers have good target tracking effect in the first 50 frames of the video. When the video exceeds 50 frames, they all fail. The reason is the small target, i.e., the basketball, is very close to the background and moves very fast, which makes trackers unable to adapt to the change of the target and mistakenly track the background as the target. Our tracker can successfully track the basketball in the first two thirds of the video, which shows that our tracker has more advantages in tracking small fast moving targets. However, in the last third of the video (from around frame 0110), when the pitcher throws the basketball, all the trackers fail to track and our tracker takes the pitcher's hand as the predicted target.

Figure 10 shows the tracking results of different trackers on the coke video sequence. The coke video shows a rotating target, i.e., a coke can, with background clutter. The difficulties of tracking this video sequence come from the target rotation and the background interference. From the tracking results, it can be seen that compared with the first two video sequences, all the target trackers dealing with this video have a relatively good tracking effect. CNHT tracker has the weakest robustness for target rotation and gets the worst tracking result. Among them, the MHT tracker and ours have the best tracking effect, always closest to the center of the ground truth. From the tracking results of the *coke* sequence, we can see that our tracker has high robustness against background interference and target rotation.

Figure 11 shows the tracking results of the paper video sequence. The paper video sequences provide scenes with camouflaged targets. The difficulty of tracking this video sequence is target rotation and similar background interference. As shown in Figure 11, DeepHKCF and CNHT both have the result of short-term target loss, which shows that the two trackers have poor robustness to the scene, where target rotation and similar background exist simultaneously. Our method performs better than the MHT method when the paper rotates. For the paper sequence, our method achieves the best tracking results, whose target box is very close to the ground truth. This is because our method considers both the local and global spectral features in the scene to overcome the interference from the background.

Figure 12 shows the tracking results of the drive video sequence. The drive video shows the translation and rotation of the target with background clutter. The difficulties of tracking the drive sequence are the target rotation deformation and the background interference. From the tracking results, it can be seen that CNHT lost the target in the early stage of the video sequence, and whose tracking ability to the target deformation caused by the target rotation is poor. In all tracking methods, the target box of our method is always closest to the ground truth, which means our method is the most accurate for target tracking. The results of the drive video sequence demonstrate that our method has strong processing ability for target deformation and background interference.

Figure 13 shows the tracking results of the kangaroo video sequence. The kangaroo video shows the tracking of a selected target in a group of fast moving kangaroos. The difficulty of tracking the video sequence is the identification of similar targets and the target's fast movement. For the kangaroo sequences, the CNHT method fails. Due to considering the local spectral features, MHT shows better identification ability than DeepHKCF. The target box of our method shows closer to the ground-truth center than MHT and achieves the best tracking effect. This demonstrates that our method has high robustness to the rapid moving target and the background interference.

From the display results of six videos, the visual comparison results of four hyperspectral trackers are consistent with the quantitative conclusions. The CNHT method has the worst tracking result. Our method and MHT have better tracking accuracy than the other two methods. Our method is the most accurate and has good robustness to the background clutter and camouflaged objects, since more spectral information is considered in building feature image. Benefiting from the channel selection strategy, our method has more advantages than MHT in tracking fast moving small targets.

4. Discussion

For hyperspectral object tracking, we propose a fast tracking method via channel selection. It is our initial goal to improve the tracking speed and make the tracking more practical. First, we design a new channel selection strategy to evaluate the quality of the hyperspectral channels from both intra-channel and inter-channel aspects. The most valuable few channels are selected from hyperspectral video avoiding the redundancy caused by some existing hyperspectral trackers focusing on the spectral information of all channels. Compared with all channels in videos, less channels are conducive to improving tracking speed. However, the number of the selected channels that can be used without reducing the tracking accuracy is worth further discussion. In this study, we limit the number to three, taking into account both tracking speed and tracking accuracy. The experimental results show that the three channels are competent for the hyperspectral object tracking task. Moreover, the second goal of our work is to improve the tracking accuracy on the basis of high tracking speed. Therefore, we improve the classical BACF tracker by modifying feature extraction, making the BACF framework more suitable for hyperspectral data. Specifically, we combine BHOG and SSHMG to form a new feature extractor which can extract the local and global spectral information from the hyperspectral channels in addition to the HOG spatial features. Although the combination of BHOG and SSHMG can effectively improve the accuracy, it may cause some spectral information to be reused, resulting in slow speed. How to extract features more effectively is still an important problem we need to consider.

From the perspective of practicality, the channel selection strategy in our method is portable. Channel Selection + Tracker can be easily and flexibly generalized to more advanced RGB-based tracking frameworks to build excellent hyperspectral object trackers.

It is worth mentioning that there are few publicly available hyperspectral video datasets for testing object tracking. Insufficient training data can seriously interfere with the accuracy and generalization of the tracking models. The “data hungry” problem causes the poor performance of deep learning based trackers. Using channel selection strategy for hyperspectral object tracking is an effective alternative and a promising idea.

5. Conclusions

We design a fast object tracking method for hyperspectral video. Although the proposed method is simple, it has achieved satisfactory results in hyperspectral videos. The proposed method first reduces the number of the input channels of hyperspectral video through a new channel selection strategy. Then the selected few channels are input to our improved BACF tracker. The channel selection and the improved BACF improve the tracking speed and the tracking accuracy, respectively. The experimental results show that our method outperforms the state-of-the-art hyperspectral trackers in both speed and accuracy. In the future work, we will deploy our tracking framework to FPGA system to make the tracking research more practical.

Author Contributions: Conceptualization, X.L.; methodology, Y.Z.; software, Y.Z.; validation, Y.Z. and X.L.; formal analysis, X.L.; investigation, Y.Z.; resources, X.L.; data curation, X.L. and B.W.; writing—original draft preparation, Y.Z.; writing—review and editing, X.L., B.W. and L.L.; visualization, Y.Z. and X.L.; supervision, X.L.; project administration, X.L. and B.W.; funding acquisition, X.L., L.L. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the European Union Horizon 2020-ULTRACEPT (778062), Shanghai Academy of Spaceflight Technology (SAST2022052), and the Practice and Innovation Funds for Graduate Students of Northwestern Polytechnical University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset composed of hyperspectral video data and RGB video data is obtained from <https://www.hsitracking.com/> (accessed on 10 September 2022) in this work.

Acknowledgments: Thanks are due to the anonymous reviewers for valuable suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Reddy, K.R.; Priya, K.H.; Neelima, N. Object Detection and Tracking—A Survey. In Proceedings of the International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015.
- Chavda, H.K.; Dhamecha, M. Moving Object Tracking Using PTZ Camera in Video Surveillance System. In Proceedings of the International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 1–2 August 2017.
- Wu, D.; Song, H.; Fan, C. Object Tracking in Satellite Videos Based on Improved Kernel Correlation Filter Assisted by Road Information. *Remote Sens.* **2022**, *14*, 4215. [\[CrossRef\]](#)
- Hu, W.; Gao, J.; Xing, J.; Zhang, C.; Maybank, S. Semi-Supervised Tensor-Based Graph Embedding Learning and Its Application to Visual Discriminant Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 172–188. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sui, Y.; Wang, G.; Zhang, L. Joint Correlation Filtering for Visual Tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 167–178. [\[CrossRef\]](#)
- Zeng, H.; Peng, N.; Yu, Z.; Gu, Z.; Liu, H.; Zhang, K. Visual Tracking Using Multi-Channel Correlation Filters. In Proceedings of the IEEE International Conference on Digital Signal Processing (DSP), Singapore, 21–24 July 2015.
- Lin, B.; Bai, Y.; Bai, B.; Li, Y. Robust Correlation Tracking for UAV with Feature Integration and Response Map Enhancement. *Remote Sens.* **2022**, *14*, 4073. [\[CrossRef\]](#)
- Xu, Y.; Zhou, X.; Chen, S.; Li, F. Deep Learning for Multiple Object Tracking: A Survey. *IET Comput. Vis.* **2019**, *13*, 355–368. [\[CrossRef\]](#)
- Liu, J.; Wang, Z.; Cheng, D.; Chen, W.; Chen, C. Marine Extended Target Tracking for Scanning Radar Data Using Correlation Filter and Bayes Filter Jointly. *Remote Sens.* **2022**, *14*, 5937. [\[CrossRef\]](#)
- Fan, J.; Song, H.; Zhang, K.; Liu, Q.; Yan, F. Real-Time Manifold Regularized Context-Aware Correlation Tracking. *Front. Comput. Sci.* **2019**, *14*, 334–348. [\[CrossRef\]](#)
- Wang, X.; Brien, M.; Xiang, C.; Xu, B.; Najjaran, H. Real-Time Visual Tracking via Robust Kernelized Correlation Filter. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017.
- Lei, J.; Liu, P.; Xie, W.; Gao, L.; Li, Y.; Du, Q. Spatial-Spectral Cross-Correlation Embedded Dual-Transfer Network for Object Tracking Using Hyperspectral Videos. *Remote Sens.* **2022**, *14*, 3512. [\[CrossRef\]](#)
- Chen, L.; Zhao, Y.; Yao, J.; Chen, J.; Li, N.; Chan, J.C.; Kong, S.G. Object Tracking in Hyperspectral-Oriented Video with Fast Spatial-Spectral Features. *Remote Sens.* **2021**, *13*, 1922. [\[CrossRef\]](#)
- Erturk, A.; Iordache, M.D.; Plaza, A. Hyperspectral Change Detection by Sparse Unmixing with Dictionary Pruning. In Proceedings of the 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, Japan, 2–5 June 2015.
- Yang, C.; Lee, W.S.; Gader, P.; Han, L. Hyperspectral Band Selection Using Kullback-Leibler Divergence for Blueberry Fruit Detection. In Proceedings of the 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Gainesville, FL, USA, 26–28 June 2013.
- Liu, Z.; Zhong, Y.; Wang, X.; Shu, M.; Zhang, L. Unsupervised Deep Hyperspectral Video Target Tracking and High Spectral-Spatial-Temporal Resolution (H^3) Benchmark Dataset. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [\[CrossRef\]](#)
- Duran, O.; Onasoglou, E.; Petrou, M. Fusion of Kalman Filter and Anomaly Detection for Multispectral and Hyperspectral Target Tracking. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Cape Town, South Africa, 12–17 July 2009.
- Xiong, F.; Zhou, J.; Chanussot, J.; Qian, Y. Dynamic Material-Aware Object Tracking in Hyperspectral Videos. In Proceedings of the 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 September 2019.
- Banerjee, A.; Burlina, P.; Broadwater, J. Hyperspectral Video for Illumination-Invariant Tracking. In Proceedings of the 1st Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Grenoble, France, 26–28 August 2009.
- Van Nguyen, H.; Banerjee, A.; Chellappa, R. Tracking via Object Reflectance Using a Hyperspectral Video Camera. In Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010.

21. Kandylakis, Z.; Karantzas, K.; Doulamis, A.; Doulamis, N. Multiple Object Tracking with Background Estimation in Hyperspectral Video Sequences. In Proceedings of the 7th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, Japan, 2–5 June 2015.
22. Qian, K.; Zhou, J.; Xiong, F.; Zhou, H.; Du, J. Object Tracking in Hyperspectral Videos with Convolutional Features and Kernelized Correlation Filter. In Proceedings of the International Conference on Software Maintenance (ICSM), Madrid, Spain, 25–27 September 2018.
23. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed Tracking with Kernelized Correlation Filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005.
25. UzKent, B.; Rangnekar, A.; Hoffman, M.J. Tracking in Aerial Hyperspectral Videos Using Deep Kernelized Correlation Filters. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 449–461. [\[CrossRef\]](#)
26. Xiong, F.; Zhou, J.; Qian, Y. Material Based Object Tracking in Hyperspectral Videos. *IEEE Trans. Image Process.* **2020**, *29*, 3719–3733. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Galoogahi, H.K.; Fagg, A.; Lucey, S. Learning Background-Aware Correlation Filters for Visual Tracking. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
28. Li, Z.; Xiong, F.; Zhou, Z.; Wang, J.; Lu, J.; Qian, Y. BAE-Net: A Band Attention Aware Ensemble Network for Hyperspectral Object Tracking. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020.
29. Liu, Z.; Wang, X.; Shu, M.; Li, G.; Sun, C.; Liu, Z.; Zhong, Y. An Anchor-Free Siamese Target Tracking Network for Hyperspectral Video. In Proceedings of the 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 March 2021.
30. Liu, Z.; Wang, X.; Zhong, Y.; Shu, M.; Sun, C. SaimHYPER: Learning A Hyperspectral Object Tracker from an RGB-Based Tracker. *IEEE Trans. Image Process.* **2022**, *31*, 7116–7129. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Li, Z.; Ye, X.; Xiong, F.; Lu, J.; Zhou, J.; Qian, Y. Spectral-Spatial-Temporal Attention Network for Hyperspectral Tracking. In Proceedings of the 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 March 2021.
32. Zhao, C.; Liu, H.; Su, N.; Yan, Y. TFTN: A Transformer-Based Fusion Tracking Framework of Hyperspectral and RGB. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [\[CrossRef\]](#)
33. Hyperspectral Object Tracking Challenge 2022. Available online: <https://www.hsitracking.com> (accessed on 10 September 2022).
34. Zhang, Y.; Li, X.; Wang, F.; Wei, B.; Li, L. A Fast Hyperspectral Object Tracking Method Based on Channel Selection Strategy. In Proceedings of the 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Rome, Italy, 13–16 September 2022.
35. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [\[CrossRef\]](#)
36. Wu, Y.; Yang, M.H. Object Tracking Benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1834–1848. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Danelljan, M.; Hager, G.; Khan, F.S.; Felsberg, M. Accurate Scale Estimation for Robust Visual Tracking. In Proceedings of the British Machine Vision Conference (BMVC), Nottingham, UK, 1–5 September 2014.
38. Danelljan, M.; Robinson, A.; Khan, F.; Felsberg, M. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016.
39. Valmadre, J.; Bertinetto, L.; Henriques, J.F.; Vedaldi, A.; Torr, P.H.S. End-to-End Representation Learning for Correlation Filter Based Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.