# **Traffic Object Tracking Based on Satellite Videos**

Subjects: Transportation Science & Technology Contributor: Zhaoxiang Zhang, Chenghang Wang, Jianing Song, Yuelei Xu

Object tracking is a hot topic in computer vision and remote sensing, and it typically employs a bounding box that locks onto the region of interest (ROI) when only an initial state of the target (in a video frame) is available. Thanks to the development of satellite imaging technology, various satellites with advanced onboard cameras have been launched to obtain very high resolution (VHR) satellite videos for military and civilian applications. Compared to traditional target tracking methods, satellite video target tracking is more efficient in motion analysis and object surveillance, and has shown great potential applications in spying on enemies, monitoring and protecting sea ice, fighting wildfires, and monitoring city trafficking, which traditional target tracking cannot even approach. The satellite video based traffic object tracking methods are reviewed under two headings, which are online tracking methods and offline tracking methods. Furthermore, the online tracking methods are grouped into four broad types: correlation filter (CF)-based, tracking-by-detection (TBD), deep learning (DL)-based, and optical flow-based methods. Finally, discussions on reviewed traffic tracking methods are delivered.

Keywords: satellite video ; traffic target tracking ; Object Tracking

# 1. Online Tracking Methods

Tracking using satellite video has confronted many challenges compared with traditional object tracking tasks because of the characteristics of satellite video data, such as large scene size, small target size, few features, and similar background. Thus, various tracking architectures have been proposed to deal with the above challenges. The mainstream solutions (**Figure 1**) to satellite video-based tracking consist of the optical flow-based method, the correlation filter (CF)-based method, the deep learning (DL)-based method, and the tracking-by-detection (TBD)-based method.



**Figure 1.** General architecture sketch of online tracking methods for traffic objects. (a) CF-based tracking methods; (b) TBD methods; (c) DL-based methods; (d) optical flow-based methods.

#### 1.1. Correlation Filter-Based Tracking Methods

The CF has yielded promising results in optical tracking tasks and is one of the most popular tracking algorithms in satellite videos. However, the CF-based tracker achieves poor results because the size of each target compared with the

entire image is too small. Several improved strategies are proposed herein for taking advantage of the CF to gain a better tracking performance. **Table 1** summarizes specific CF-based tracking methods. As shown in **Table 1**, recent CF-based tracking methods for traffic objects are of three kinds: (1) kernel correlation tracker (KCF) with multi-frame case; (2) KCF for target motion case; and (3) KCF aided by kernel adaptation. The general pipeline of the CF-based tracking methods is depicted in **Figure 1**a.

Methods	Ref.	Year	Description
KCF + multi-frames	[1]	2017	KCF with the three-frame-difference
	[2]	2022	KCF with Multi-feature fusion
	[3]	2019	Improved discriminative CF for small objects tracking
	[ <u>4</u> ]	2019	High-speed CF-based tracker for object tracking
KCF + target motion	[5]	2019	KCF embedded with motion estimations
	[6]	2021	Feature fusion, position compensation, local object region
	[7]	2018	VCF using velocity feature and inertia mechanism
	[8]	2019	Hybrid KCF with histogram of oriented gradient
	<u>[9]</u>	2021	Rotation-adaptive CF
KCF + kernel adaptation	[ <u>10]</u>	2022	Rotation-adaptive CF with motion constraint
	[ <u>11]</u>	2022	Spatial-Temporal regularized CF with interacting multiple model
	[ <u>12]</u>	2022	Kernelized CF with color-name features

#### Table 1. Summary of various CF-based traffic objects tracking methods.

In 2017, Ref. <sup>[1]</sup> presented a new object tracking method by taking advantage of the KCF and the three-frame-difference method to deal with satellite videos. The integrated model combined the shape information provided by the KCF tracker and the change information from the three-frame-difference method into the final tracking results. Three videos that described the conditions of Canada, Dubai, and New Delhi were introduced, with the target of moving trains and cars. The image sizes of these three videos were 3840 ×2160 pixels for both the first and second videos and 3600 × 2700 pixels for the third one. The average center location error (CLE) and the average overlap score were 11 pixels and 71%, respectively. Later, Ref. <sup>[2]</sup> presented a KCF embedded method that fused multi-feature fusion and compensates motion trajectory to track fast-moving objects in satellite videos. The contributions of the suggested algorithm were multifold. First, a multi-feature fusion strategy was proposed to describe an object comprehensively, which was challenging for the single-feature approach. Second, a subpixel positioning method was developed to calculate accurate object localization that was further used to improve the tracking accuracy. Third, the adaptive Kalman filter (AKF) was introduced to compensate for the KCF tracker results and reduce the object's bounding box drift, solving the moving object occlusion problem. Compared to the KCF algorithm, the algorithm improved the tracking accuracy and the success rate with over 17% and 18% on average.

In 2019, Ref. <sup>[3]</sup> developed an improved discriminative CF for small object tracking in satellite videos. Instead of employing a change detection tracking model, the authors first proposed a spatial mask to promote the CF to give different contributions depending on the spatial distance. The Kalman filter (KF) was then applied to predict the target position in the large and analogous background region. Next, the integrated strategy was applied to combine the improved CF tracker and pose estimation algorithm. The proposed model was implemented on the Chang Guang Satellite dataset with an image resolution of 3840 × 2160 pixels. The authors calculated success rate, precision, and frame per second (FPS) measurement indicators to evaluate the performance, achieving the result of 0.725, 0.96, and 1500, respectively. Comparing with other video tracking methods, including Channel and Spatial Reliability Tracker (CSRT) <sup>[13]</sup>, Efficient Convolution Operator Tracker (ECOT) <sup>[14]</sup>, long-term correlation tracker <sup>[15]</sup>, and KCF models, the proposed method performed best.

Later, a high-speed CF-based tracker was derived by <sup>[ $\Delta$ ]</sup> for object tracking in satellite videos. The authors introduced the global motion characteristics of the moving vehicle target to constrain the tracking process. By integrating the position and velocity KF, the trajectory of the moving target was corrected. The tracking confidence module (TCM) was proposed to couple the KF and CF algorithms tightly, in which the confidence map of the tracking results was obtained by the CF and passed to the KF for a better prediction. The authors cropped the satellite videos of SkySat-1 and Jilin-1 into nine short

sequences, which contained 31 moving objects in total, and then applied their method to the cropped satellite videos. Five metrics, namely, expected average overlap (EAO), accuracy, robustness, average overlap, and FPS, were used to evaluate the capability of the proposed method for object tracking, with the results of 0.7205, 0.71, 0.00, 0.7053, and 1094.67, respectively. Thus, the introduced technique was verified to be effective and fast for real-time vehicle tracking in satellite videos. Similarly, Ref. <sup>[5]</sup> studied a KCF embedded with motion estimations to track satellite video targets. The authors developed an innovative motion estimation (ME) algorithm combining the KF and motion trajectory to average and mitigate the boundary effects of KCF. An integrated strategy based on motion estimation was proposed to solve the problem of tracking failure when a moving object was partially or completely occluded. The experimental dataset consisted of 11 videos with a resolution of 1 m from the Jilin-1 satellite constellation. The area under curve (AUC), CLE, overlap score, and FPS measurement indicators were utilized to evaluate the tracking performance, which is 72.9, 94.3, 96.4, and 123, respectively. Compared with other object tracking methods, the developed model gained the best results. Furthermore, Ref. <sup>[6]</sup> proposed an improved KCF to track the object in satellite videos. The improvements of the proposed algorithm were: (1) fusing the different features of the object, (2) proposing a motion position compensation algorithm by combining the KF and motion trajectory, and (3) extracting the local object region for normalized cross-correlation matching. Thus, the algorithm was able to track the moving object in satellite video with high accuracy effectively.

Differing from the above feature-kernel-based tracking methods, Ref. [2] considered the extremely inadequate quality of target features in satellite videos. The authors designed a velocity correlation filter (VCF) by employing the velocity feature and inertia mechanism to construct a KCF for satellite video target tracking. The velocity feature, with the high discriminative ability and inertial mechanism, could help to detect moving targets and prevent model drift in satellite videos. The experiment results showed that the AUC scores in precision and success plots of the proposed method reached 0.941 and 0.802, respectively. Moreover, the presented tracker had a favorable speed compared to other stateof-the-art methods, running at over 100 FPS. Later, Ref. [8] designed a hybrid kernel correlation filter tracker for satellite video tracking. This approach integrated the optical flow features with the histogram of oriented gradient and obtained competitive results. Similarly, Ref. [9] presented a rotation-adaptive CF tracking algorithm to address the problem caused by the rotation objects. The authors proposed an object rotation estimation method to keep the feature map stable for the object rotation and achieved the capability of estimating the change in the bounding box size. Ref. [10] decoupled the rotation and translation motion patterns and developed a novel rotation adaptive tracker with motion constraints. Experiments based on the Jilin-1 satellite dataset and International Space Station dataset demonstrated the superiority of the proposed method. To handle the occlusion problem during the satellite tracking, Ref. <sup>[11]</sup> developed a spatial-temporal regularized correlation filter algorithm with interacting multiple models. The authors utilized the interacting multiple models to predict the target position when the target is occluded. Similarly, Ref. <sup>[12]</sup> designed a kernelized correlation filter based on the color-name feature and Kalman prediction. Experiment results on Jilin-1 datasets show that the proposed algorithm has stronger robustness for several complex situations such as rapid target motion and similar object interference.

#### 1.2. TBD Methods

Detection association strategy in computer vision is one of the popular methods for multi-target tracking <sup>[16]</sup>. By assigning detected candidates of each frame into trackers, the motion interpolation is utilized to retrieve the short-term missing detected candidates. This type of tracker is anointed TBD. However, unique characteristics of satellite videos, including low frame rate, less discriminative appearance information, and lacking color features, bring further challenges to current TBD methods. **Table 2** summarizes specific references of the TBD methods and **Figure 1**b shows the general pipeline of this kind of method, in which four types are further divided on the basis of the tracking features. These are motion feature-based, hyperspectral image-based, graph-based, and discriminative-based TBD methods.

Method	Ref.	Year	Description
Motion footure based tracking	[ <u>17]</u>	2017	Local context tracker
Motion reature-based tracking	[ <u>18]</u>	2021	SFMFT for multiple moving objects
	[ <u>19]</u>	2016	Study hyperspectral and spatial domain information
Hyperspectral image-based tracking	[20]	2017	Real-time HLT method
Graph-based tracking	[ <u>21]</u>	2010	Unified relation graph approach from road structure

 Table 2. Summary of TBD methods for traffic objects tracking.

Method	Ref.	Year	Description
	[22]	2017	Bayesian classification with motion smoothness constraint
Discriminative-based tracking	[ <u>23]</u>	2019	Multi-morphological cue based discrimination strategy
	[ <u>24]</u>	2020	TBD with filter training mechanism

In the tracking by detection method, the detection models play an important role in enhancing the tracking performance. Classic detectors such as YOLO <sup>[25]</sup>, CenterNet <sup>[26]</sup>, and CornerNet <sup>[27]</sup> have been applied for object tracking. For example, Ref. <sup>[28]</sup> unified Cornernet and data association to achieve a better speed-accuracy trade-off for multi-object tracking while eliminating the extra feature extraction process.

To reduce the dependency on motion detection of frame differencing and appearance information, Ref. <sup>[12]</sup> introduced a local context tracker. In their method, the local context tracker explored spatial relations for the target to avoid unreasonable model deformation in the next frame. The merged detection results in the detection association were explicitly handled, and short tracks were produced by associating hypotheses. The track association fused the results from two trackers and updated the "track pool" to improve the tracking performance. The designed model was tested on WPAFB Sequence and Rochester Sequence containing 410 and 44 tracks. Multiple metrics, namely, Recall, Precision, and the number of breaks per track (B/T), were introduced to analyze the performance, with the results of 0.606, 0.99, and 0.159, showing that the proposed method outperformed the state-of-the-art methods in satellite video-based tracking.

Aiming at tracking multiple moving objects, Ref. <sup>[18]</sup> proposed the slow feature and motion feature-guided multi-object tracking (SFMFT) method by using the slow features and motion features. Specifically, the authors developed a nonmaximum suppression (NMS) module to assist the object detection by utilizing the sensitivity of slow feature analysis to the changed pixels. This method reduced the amount of static false alarms and supplemented missed objects, further improving the recall rate by increasing the confidence score of the correctly detected object bounding boxes. The superiority of the proposed method was evaluated and demonstrated with three satellite videos.

On the other hand, Ref. <sup>[19]</sup> presented a real-time tracking method that exploits the hyperspectral and spatial domain information, aiming to reduce false alarm tracking rates. In their method, the individual feature map was computed for each hyperspectral band and then fed to an adaptive fusion method. Therefore, the fusion map with reduced noise could help to detect the targets from the background pixels efficiently. The CLIFF-2007 dataset with 0.3 cm Ground Sampling Distance (GSD) and 50 tracking targets was used to evaluate the suggested techniques. In terms of the track purity and target purity, the proposed hyperspectral feature-based method outperformed the Red-Green-Blue (RGB) only features, with the results of 64.37 and 57.49, respectively. Compared with their previous work, Ref. <sup>[20]</sup> designed an improved real-time hyperspectral likelihood maps-aided tracking (HLT) method. An online generative target model is proposed and revised for the tracking system of a target detection segment, considering the hyperspectral channels ranging from visible to infrared wavelengths. An adaptive fusion method is proposed to combine likelihood maps from multiple bands of hyperspectral imagery into one single more distinctive representation. The experimental outcomes indicate that the proposed model is able to track the traffic targets accurately.

Instead of exploring the tracking features from the targets, Ref. <sup>[21]</sup> developed a unified relation graph approach to explore vehicle behavior models from road structure and regulate object-based vertex matching in multi-vehicle satellite videos. The proposed vehicle travel behavior models generated additional constraints for better matching scores. Moreover, the authors utilized three-frame moving object detection to initialize vehicle tracks and a tracking-based target indicator to reduce miss-detection and refine the target location. The dataset used for evaluation was collected by a single camera covering a 1 km<sup>2</sup> area with a frame rate of 1 Hz. The Multiple Object Tracking Accuracy (MOTA) <sup>[29]</sup> was introduced as a metric for the accuracy assessment and was 0.85 achieved by the proposed method, thereby indicating satisfactory results for satellite video tracking. The model could be further improved by preparing extra high-quality satellite videos with tracking labels.

The above-discussed methods can be seen as graph-based methods, which explore the target movement model according to their graph features. There is another well-studied strategy for object tracking based on the discriminative method. In 2017, Ref. <sup>[22]</sup> proposed a Bayesian classification considering the motion smoothness constraint to track vehicles in satellite videos. The authors introduced the gray level similarity feature to describe the likelihood of the target with the assumption of motion smoothness, and the posterior probability was used to identify the tracking target position. Additionally, a KF was introduced to enhance the robustness of tracking processing. The SkySat and Jilin-1 satellite dataset were applied to evaluate the proposed model, showing the superiority and potential of the model for object

tracking from remote sensing imagery. Later, Ref. <sup>[23]</sup> presented a modified detection-tracking framework to identify and track small moving vehicles in satellite sequences. An original detection algorithm was developed based on local noise modeling and exponential probability distribution. After detection, a discrimination strategy based on the multi-morphological cue was designed to further identify correct vehicle targets from noises. The suggested method was employed in the Chang Guang Satellite dataset. F1 score, recall, precision, Jaccard Similarity, MOTA, and Multiple Object Tracking Precision (MOTP) were calculated to assess classification performance, with the results of 0.71, 63.06, 81.04, 0.55, 0.46, and 0.52, respectively. Furthermore, Ref. <sup>[24]</sup> exploited the circulant structure of TBD with Kernels, and established a filter training mechanism for the target and background to improve the discrimination ability of the tracking algorithm. Tracking experiments with nine sets of Jilin-1 satellite videos showed competitive performance with targets under weak feature attributes.

#### 1.3. DL-Based Tracking Methods

CNN models have achieved significant success in many vision tasks, which inspires researchers to explore their capabilities in tracking problems. State-of-the-art CNN-based trackers have made remarkable progress toward this goal <sup>[14][30][31][32]</sup>, showing more robust than traditional methods with a large training dataset. However, DL-based trackers need to adapt to satellite videos due to the challenges of the satellite videos-based target tracking issues. **Figure 1**c illustrates the general pipeline of DL-based methods, where the DL modules are utilized in the Siamese architecture to extract the appearance features. Moreover, the DL modules can be introduced into CF-based methods and TBD methods, running as the feature extractor and feature detector. For instance, Ref. <sup>[33]</sup> utilized CNN to extract hyperspectral domain features and a kernel-based CF dealing with the satellite video tracking problem.

A SN is an CNN-based approach that applies the same weights while working in tandem on two different input vectors to compute comparable output vectors, which is typically utilized for comparing similar instances in different type sets. Thus, it is a natural idea to apply the SN in the object tracking task  $\frac{32}{2}$ . In 2019, Ref.  $\frac{34}{2}$  constructed a fully convolutional SN with shallow-layer features to retrieve fine-grained appearance features for space-borne satellite video tracking. Predicting attention combined Gaussian Mixture Model (GMM), and KF was utilized to deal with tracking target occlusion and the obscure problem. The proposed method was validated by three high-resolution satellite videos quantitatively, which outperformed the state-of-the-art tracking methods with an FPS of 54.83. Similarly, a deep Siamese network (DSN) incorporating an interframe difference centroid inertia motion (ID-CIM) model was proposed in Ref. [35], in which the ID-CIM mechanism was proposed to alleviate model drift. The DSN inherently included a template branch and a search branch and extracted the features from these two branches. A Siamese region proposal network was then employed to obtain the target position in the search branch. Meanwhile, <sup>[36]</sup> investigated a lightweight parallel network with a high spatial resolution to locate the small objects in satellite videos, namely, the Hign-resolution Siamese network (HRSiam). A pixel-level refining model based on online moving object detection and adaptive fusion was proposed to enhance the tracking robustness in satellite videos. By modeling the video sequence in time, the HRSiam detected the moving targets in pixels with the advantage of tracking and detecting. The authors reported that their proposed HRSiam achieved stateof-the-art tracking performance while running at over 30 FPS.

Recently, RNs have been studied and shown promising performance in the field of satellite video-based object tracking. For example, Ref. <sup>[37]</sup> introduced a convolutional regression network with appearance and motion feature (CRAM), which consisted of training and tracking two phases. In the training phase, the two RNs were trained with different appearance and motion features respectively. In the tracking phase, the model responses were weighted by their qualities measured from the peak-to-sidelobe ratio (PSR) and then integrated for the final target location prediction <sup>[38]</sup>. To evaluate the performance of the proposed network, the authors collected nine small sequences with a total number of 31 moving vehicles, which were cropped from the SkySat-1 and Jilin-1 satellite videos. The average overlap measure and expected average overlap indices were analyzed, with the results of 0.7 and 0.7286, thereby demonstrating the efficiency of the presented network in object tracking from high-resolution remote sensing videos. Later, Ref. <sup>[39]</sup> suggested a cross-frame keypoint-based detection network based on a two-branch Long short-term memory (LSTM). The spatial information and motion information of moving targets are extracted for better tracking of the missed or occluded vehicles. Experimental results on Jilin-1 and SkySat satellite videos illustrated the effectiveness of the proposed tracking algorithms.

Furthermore, a prediction network (PN) was studied by <sup>[40]</sup>, which predicted the location probability of the target in each pixel in the next frame using the fully convolutional network (FCN) learned from previous results. The authors further introduced a segmentation method to generate the feasible region with an assigned high probability for the target in each frame. Experiments were carried out with nine satellite videos taken from the JiLin-1, indicating the superiority of the proposed method, as the author reported.

By taking advantage of both the SN and RN, Ref. <sup>[41]</sup> proposed a two-stream deep neural network (SRN) that combined a SN and a motion RN for satellite object tracking. In Ref. <sup>[41]</sup>, a trajectory fitting motion model (FTM) based on history trajectories was employed to further alleviate model drift. Comprehensive experiments demonstrated that their method performed favorably compared with the state-of-the-art tracking methods. Additionally, by exploring the temporal and spatial context, the object appearance model, and the motion vector from occluded targets, Ref. <sup>[42]</sup> designed a Reinforcement learning (RL)-based approach to enhance the tracking performance under complete occlusion. In addition, Ref. <sup>[43]</sup> explored the potential of graph convolution (GC) for multi-object tracking and modeled the satellite video tracking as a graph information reasoning procedure from the multitask learning perspective. Compared with state-of-the-art multi-object trackers, the tracking accuracy of this model increased by 20%.

To sum up, **Table 3** illustrates recent published articles that study the DL-based tracking methods. As listed in **Table 3**, the SN-based models are widely utilized for object tracking in the remote sensing area. The CNN combined with CF tracking is another popular trend, which integrates the efficiency of the CF method and robustness of the CNN. Meanwhile, due to the advantages in time-series image processing, RN-based approaches have shown their potential for advanced tasks, such as long-term tracking or tracking with occlusion.

Method	Ref.	Year	Description
	<u>[34]</u>	2019	Predicting attention-inspired SN
SN-based tracking	[35]	2021	DSN + ID-CIM
	<u>[36]</u>	2021	Lightweight parallel network with a high spatial resolution
CNN combined with CF	[33]	2018	Kernelized CF utilizing deep CNN features
RN-based tracking	[ <u>37]</u>	2020	CRAM, RN-based training
RN-based tracking	[ <u>39]</u>	2021	A two-branch LSTM
PN-based tracking	<u>[40]</u>	2021	PN to predict the location probability of the target
Combined SN and RN	[ <u>41]</u>	2022	SRN followed by FTM
RL-based tracking	[ <u>42]</u>	2022	RL to track objects under occlusion
GC-based tracking	[43]	2022	Tracking via GC-based multitask reasoning

#### Table 3. Summary of DL-based traffic objects tracking methods.

#### 1.4. Optical Flow-Based Methods

The optical flow method utilizes the apparent motion of the brightness patterns in the image to detect moving objects. The algorithm output can provide vital information for the tiny movements of an object <sup>[44]</sup>. It is worth noting that the background relative to the interested target is generally constant in satellite videos. Therefore, the image target and background can be separated by optical flow efficiently. If target objects move too slow to be analyzed with optical flow, multi-frame differences can be employed to improve the tracking performance <sup>[45]</sup>. **Table 4** summarizes typical methods of the optical flow-based methods. Global feature-based optical flow is an old-fashioned method of tracking objects from remote sensing images, whereas local feature-based optical flow methods are gaining popularity recently. The general architecture of the optical flow based methods is depicted in **Figure 1**d.

Table 4. Summary of optical flow based traffic objects' tracking methods.

Method	Ref.	Year	Description
Global feature-based	[ <u>46</u> ]	2013	Three-frame differencing scheme
	[47]	2019	Multi-frame optical flow tracker
Local feature-based	[48]	2021	SLIC + optical flow
	[49]	2022	HoG + optical flow

Earlier researchers utilized a three-frame differencing scheme to detect and track vehicles globally. <sup>[46]</sup>. In Ref. <sup>[46]</sup>, the authors firstly proposed a box filter to reduce the seam artifacts caused by considerable radiometric changes in different focal planes of the original stitched image. The grid was chosen such that tiles were approximately 1000 × 1000 pixels.

The tile processors then enabled the global parallelism necessary to achieve real-time performance. In addition, the tile patches were further set up to overlap by about 80 pixels at each border to ensure that vehicles near the edges are included.

More recently, local feature-based methods were developed, and Ref. <sup>[42]</sup> implemented a multi-frame optical flow tracker to track the vehicles in satellite videos. The author first proposed a Lucas–Kanade optical flow method to obtain the optical flow field. The Hue-Saturation-Value (HSV) color system was then utilized to convert the two-dimensional optical flow field into a three-bands color image. Finally, the integral image was adapted to obtain the most probable position of the target. Five satellite videos provided by UrtheCast Corp. and Chang Guang Satellite Technology Co., Ltd. were applied in experiments, showing that the proposed method can track slightly moving objects accurately. Additionally, an optical flow motion estimation combined with a superpixel algorithm was presented by <sup>[48]</sup>. The authors used simple linear iterative clustering (SLIC) to realize superpixels, which made the object a more regular and compact shape. The output of the superpixel algorithm was then fed to the optical flow method to obtain and label the moving object. In 2022, Ref. <sup>[49]</sup> fused the histogram of oriented gradient (HoG) features and optical flow features to enhance the representation information of the targets. The author also developed a disruptor-aware mechanism to weaken the influence of background noise. Experimental results show that the proposed algorithm achieves high tracking performance with target occlusion.

### 2. Offline Tracking Methods

Online tracking can only use existing frames for tracking model updates, whereas offline tracking methods can benefit from all keyframes providing the smoothness constraint <sup>[50]</sup>. Since the satellite videos are generally downloaded from the aerial platform in advance, the offline video tracking models are implemented to entire video frames. Compared to the online video tracking algorithms, offline tracking is typically formulated as a global optimization problem to obtain the global optimal tracks. Furthermore, hyperspectral videos are usually introduced to improve the performance of the offline tracking models. **Table 5** summarizes the reviewed offline traffic object tracking methods, which are divided depending on how many steps to obtain the tracking result. One-step-based methods utilize the tracker only, while two-step algorithms consist of both detector and tracker.

Method	Ref.	Year	Description
One step-based	<u>[51]</u>	2014	Two paralleled trackers for initialization and tracking
	[52]	2021	3D variation regularization + PCA
	[53]	2018	Global data association approach
Two step-based	<u>[54]</u>	2019	DTS for traffic parameters estimating
	<u>[55]</u>	2019	DTS for vehicle tracking

Table 5. Summary of offline traffic object tracking methods.

In 2014, Ref. <sup>[51]</sup> proposed a fused framework for tracking multiple cars from satellite videos, in which two trackers worked in parallel. One tracker provided target initialization and reacquisition through detections from background subtraction. The other offered a frame to frame tracking by a target state regressor. A sequence from a publicly available wide-area aerial imagery dataset WPAFB was applied to test the proposed framework. Tracking metric indicators, namely, track swaps, track breaks, and overall MOTA, were calculated with 0.20, 0.92, and 0.41, respectively, in terms of detection and tracking metrics. Later, Ref. <sup>[52]</sup> incorporated a three-dimensional (3D) total variation regularization into the robust PCA model, in order to extract the moving targets from the background. Evaluation results on real remote sensing videos have demonstrated the advantage of this approach.

An offline two-step global data association approach was later presented in Ref. <sup>[53]</sup> to track multiple targets using satellite videos. The authors extended the spatial grid flow model to cover the possible connectivities in a wider temporal neighboring, making sure the association matches temporal-unlinked detections. Then, a KF-based tracklet transition probability was customized to link tracklets within large temporal intervals. To demonstrate traffic tracking capabilities, the proposed method was evaluated on a dataset that was cropped from a satellite high definition video captured by SkySat-1 on 25 March 2014.

On the other hand, Ref. <sup>[54]</sup> contributed to the integration of the two-step offline tracking algorithm, developing a complete and effective offline detection-tracking system (DTS) using satellite videos to estimate traffic parameters. In their system,

a video preprocessing step is firstly applied to obtain the background. The moving targets were then checked over time to construct the target trajectories. A threshold method based on target displacement and velocity was utilized to eliminate false positives. A satellite video captured over Las Vegas from the SkySat-1 satellite with 30 FPS was applied to the proposed method. The results still revealed the limitation of the said method which was the inability of noise removal conditions to filter out tall buildings' relief displacement. Meanwhile, Ref. <sup>[55]</sup> offered an efficient DTS to track vehicles in multi-temporal remote sensing images. In the detection phase, the authors applied background subtraction, reduced searching space, and combined road prior information to improve detection accuracy. In the tracking phase, a dynamic association method under state judgment rules was designed to associate all potential target candidates. Additionally, a group dividing method was proposed to further improve the tracking accuracy. The proposed model was evaluated on a remote sensing video dataset with a 10 FPS frame rate and 4096 × 2160 pixels resolution. Completeness, Correctness, and Quality indices were utilized for the performance assessment with the results of 0.99, 0.97, and 0.97, showing the effectiveness of the presented method in tracking small vehicles from satellite sequences.

## Abbreviations

3D	three-dimensional
AKF	adaptive Kalman filter
AHI	advanced Himawari imager
AUC	area under curve
B/T	breaks per track
CLE	center location error
CF	correlation filter
CNN	convolutional neural network
CRAM	convolutional regression network with appearance and motion feature
CSRT	Channel and Spatial Reliability Tracker
СVН	Canada Vancouver harbor
DL	deep learning
DMSM	dynamic multiscale saliency map
DS	Difficulty Score
DSN	deep Siamese network
DTS	detection-tracking system
EAO	expected average overlap
ECOT	Efficient Convolution Operator Tracker
FCL	fully connected layer
FCN	fully convolutional network
FTM	fitting motion model
FPS	frame per second
GC	graph convolution
GMM	Gaussian Mixture Model
GICA	Geometrical Independent Component Analysis
GPS	Global Positioning System
GNN	global nearest neighbor
GRU	gated recurrent unit
GSD	Ground Sampling Distance
HRSiam	High-resolution Siamese network

HCF	Hierarchical Convolutional Features
HLT	hyperspectral likelihood maps-aided tracking
HoG	histogram of oriented gradient
HSV	Hue-Saturation-Value
ID-CIM	interframe difference centroid inertia motion
ISS	International Space Station
ICP	iterative closest point
JTWC	Joint Typhoon Warning Center
KF	Kalman filter
KCF	kernel correlation tracker
LEO	Low Earth Orbiting
LSTM	Long short-term memory
MLTB	multi-level tracking benchmark
МСС	maximum cross correlation
MNN	matrix neural network
MDDCM	multiscale dual-neighbor difference contrast measure
ML	machine learning
ME	motion estimation
ΜΟΤΑ	Multiple Object Tracking Accuracy
МОТР	Multiple Object Tracking Precision
NMS	nonmaximum suppression
PCA	principal component analysis
PSR	peak-to-sidelobe ratio
PN	prediction network
RGB	Red-Green-Blue
RL	Reinforcement learning
RN	regression network
SFMFT	slow feature and motion feature-guided multi-object tracking
SLIC	simple linear iterative clustering
SRN	two-stream deep neural network
SN	Siamese network
ТСМ	tracking confidence module
TBD	tracking-by-detection
UAV	unmanned aerial vehicle
VCF	velocity correlation filter
VHR	very high resolution
WPAFB	Wright Patterson Air Force Base

#### References

- 1. Du, B.; Sun, Y.; Cai, S.; Wu, C.; Du, Q. Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm. IEEE Geosci. Remote Sens. Lett. 2017, 15, 168–172.
- 2. Liu, Y.; Liao, Y.; Lin, C.; Jia, Y.; Li, Z.; Yang, X. Object Tracking in Satellite Videos Based on Correlation Filter with Multi-Feature Fusion and Motion Trajectory Compensation. Remote Sens. 2022, 14, 777.
- Chen, X.; Sui, H. Real-time tracking in satellite videos via joint discrimination and pose estimation. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Moscow, Russia, 13–15 May 2019; pp. 23–29.
- 4. Guo, Y.; Yang, D.; Chen, Z. Object Tracking on Satellite Videos: A Correlation Filter-Based Tracking Method With Trajectory Correction by Kalman Filter. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2019, 12, 3538–3551.
- 5. Xuan, S.; Li, S.; Han, M.; Wan, X.; Xia, G.S. Object Tracking in Satellite Videos by Improved Correlation Filters With Motion Estimations. IEEE Trans. Geosci. Remote Sens. 2019, 58, 1074–1086.
- Yaosheng, L.; Yurong, L.; Cunbao, L.; Zhaoming, L.; Xinyan, Y.; Aidi, Z. Object Tracking in Satellite Videos Based on Improved Correlation Filters. In Proceedings of the 2021 13th International Conference on Communication Software and Networks (ICCSN), Chongqing, China, 4–7 June 2021; pp. 323–331.
- Shao, J.; Du, B.; Wu, C.; Wu, J.; Hu, R.; Li, X. VCF: Velocity correlation filter, towards space-borne satellite video tracking. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.
- Shao, J.; Du, B.; Wu, C.; Zhang, L. Can We Track Targets From Space? A Hybrid Kernel Correlation Filter Tracker for Satellite Video. IEEE Trans. Geosci. Remote Sens. 2019, 57, 8719–8731.
- 9. Xuan, S.; Li, S.; Zhao, Z.; Zhou, Z.; Zhang, W.; Tan, H.; Xia, G.; Gu, Y. Rotation adaptive correlation filter for moving object tracking in satellite videos. Neurocomputing 2021, 438, 94–106.
- 10. Chen, Y.; Tang, y.; Ha, T.; Zhang, Y.; Zou, B.; Feng, H. RAMC: A Rotation Adaptive Tracker with Motion Constraint for Satellite Video Single-Object Tracking. Remote Sens. 2022, 14, 3108.
- 11. Li, Y.; Bian, C. Object Tracking in Satellite Videos: A Spatial-Temporal Regularized Correlation Filter Tracking Method With Interacting Multiple Model. IEEE Geosci. Remote Sens. Lett. 2022, 19, 6511105.
- 12. Pei, W.; Lu, X. Moving Object Tracking in Satellite Videos by Kernelized Correlation Filter Based on Color-Name Features and Kalman Prediction. Wirel. Commun. Mob. Comput. 2022, 2022, 9735887.
- Farkhodov, K.; Lee, S.H.; Kwon, K.R. Object Tracking using CSRT Tracker and RCNN. In Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020)—Volume 2: BIOIMAGING, Valletta, Malta, 24–26 February 2020; pp. 209–212.
- Danelljan, M.; Bhat, G.; Shahbaz Khan, F.; Felsberg, M. ECO: Efficient convolution operators for tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6638–6646.
- 15. Ma, C.; Yang, X.; Zhang, C.; Yang, M.H. Long-term correlation tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA USA, 7–12 June 2015; pp. 5388–5396.
- 16. Zhao, H.; Yang, G.; Wang, D.; Lu, H. Deep mutual learning for visual object tracking. Pattern Recognit. 2021, 112, 107796.
- Chen, B.J.; Medioni, G. Exploring local context for multi-target tracking in wide area aerial surveillance. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; pp. 787–796.
- 18. Wu, J.; Su, X.; Yuan, Q.; Shen, H.; Zhang, L. Multi-Vehicle Object Tracking in Satellite Video Enhanced by Slow Features and Motion Features. IEEE Trans. Geosci. Remote Sens. 2021, 60, 5616426.
- 19. Uzkent, B. Real-Time Aerial Vehicle Detection and Tracking Using a Multi-Modal Optical Sensor; Rochester Institute of Technology: Rochester, NY, USA, 2016.
- Uzkent, B.; Rangnekar, A.; Hoffman, M. Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 39–48.
- 21. Xiao, J.; Cheng, H.; Sawhney, H.; Han, F. Vehicle detection and tracking in wide field-of-view aerial video. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 679–684.

- 22. Wu, J.; Zhang, G.; Wang, T.; Jiang, Y. Satellite video point-target tracking in combination with motion smoothness constraint and grayscale feature. Acta Geod. Cartogr. Sin. 2017, 46, 1135–1146.
- Ao, W.; Fu, Y.; Hou, X.; Xu, F. Needles in a Haystack: Tracking City-Scale Moving Vehicles From Continuously Moving Satellite. IEEE Trans. Image Process. 2019, 29, 1944–1957.
- 24. Wang, Y.; Wang, T.; Zhang, G.; Cheng, Q.; Wu, J.q. Small target tracking in satellite videos using background compensation. IEEE Trans. Geosci. Remote Sens. 2020, 58, 7010–7021.
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision in Addition, Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 6569–6578.
- 27. Law, H.; Deng, J. CornerNet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.
- 28. Feng, X.; Wu, H.M.; Yin, Y.H.; Lan, L.B. CGTracker: Center Graph Network for One-Stage Multi-Pedestrian-Object Detection and Tracking. J. Comput. Sci. Technol. 2022, 37, 626–640.
- Kasturi, R.; Goldgof, D.; Soundararajan, P.; Manohar, V.; Garofolo, J.; Bowers, R.; Boonstra, M.; Korzhova, V.; Zhang, J. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. IEEE Trans. Pattern Anal. Mach. Intell. 2008, 31, 319–336.
- 30. Ma, C.; Huang, J.B.; Yang, X.; Yang, M.H. Hierarchical convolutional features for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 3074–3082.
- Nam, H.; Han, B. Learning multi-domain convolutional neural networks for visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4293– 4302.
- Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11– 14 October 2016; pp. 850–865.
- 33. Uzkent, B.; Rangnekar, A.; Hoffman, M.J. Tracking in aerial hyperspectral videos using deep kernelized correlation filters. IEEE Trans. Geosci. Remote Sens. 2018, 57, 449–461.
- Shao, J.; Du, B.; Wu, C.; Pingkun, Y. PASiam: Predicting Attention Inspired Siamese Network, for Space-Borne Satellite Video Tracking. In Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019; pp. 1504–1509.
- Zhu, K.; Zhang, X.; Chen, G.; Tan, X.; Liao, P.; Wu, H.; Cui, X.; Zuo, Y.; Lv, Z. Single object tracking in satellite videos: Deep Siamese network incorporating an interframe difference centroid inertia motion model. Remote Sens. 2021, 13, 1298.
- 36. Shao, J.; Du, B.; Wu, C.; Gong, M.; Liu, T. Hrsiam: High-resolution siamese network, towards space-borne satellite video tracking. IEEE Trans. Image Process. 2021, 30, 3056–3068.
- 37. Hu, Z.; Yang, D.; Zhang, K.; Chen, Z. Object Tracking in Satellite Videos Based on Convolutional Regression Network With Appearance and Motion Features. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2020, 13, 783–793.
- Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550.
- Feng, J.; Zeng, D.; Jia, X.; Zhang, X.; Li, J.; Liang, Y.; Jiao, L. Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos. ISPRS J. Photogramm. Remote Sens. 2021, 177, 116–130.
- 40. Zhang, W.; Jiao, L.; Liu, F.; Li, L.; Liu, X.; Liu, J. MBLT: Learning Motion and Background for Vehicle Tracking in Satellite Videos. IEEE Trans. Geosci. Remote Sens. 2021, 60, 4703315.
- 41. Ruan, L.; Guo, Y.; Yang, D.; Chen, Z. Deep Siamese Network with Motion Fitting for Object Tracking in Satellite Videos. IEEE Geosci. Remote Sens. Lett. 2022, 19, 6508005.
- 42. Cui, Y.; Hou, B.; Wu, Q.; Ren, B.; Wang, S.; Jiao, L. Remote Sensing Object Tracking With Deep Reinforcement Learning Under Occlusion. IEEE Trans. Geosci. Remote Sens. 2021, 60, 5605213.

- 43. He, Q.; Sun, X.; Yan, Z.; Li, B.; Fu, K. Multi-Object Tracking in Satellite Videos With Graph-Based Multitask Modeling. IEEE Trans. Geosci. Remote Sens. 2022, 60, 5619513.
- 44. Nejadasl, F.K.; Gorte, B.G.; Hoogendoorn, S.P. Optical flow based vehicle tracking strengthened by statistical decisions. ISPRS J. Photogramm. Remote Sens. 2006, 61, 159–169.
- 45. Wu, C.; Zhang, L.; Du, B. Kernel slow feature analysis for scene change detection. IEEE Trans. Geosci. Remote Sens. 2017, 55, 2367–2384.
- 46. Keck, M.; Galup, L.; Stauffer, C. Real-time tracking of low-resolution vehicles for wide-area persistent surveillance. In Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), Clearwater Beach, FL, USA, 15–17 January 2013; pp. 441–448.
- 47. Du, B.; Cai, S.; Wu, C. Object Tracking in Satellite Videos Based on a Multiframe Optical Flow Tracker. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2019, 12, 3043–3055.
- Xu, G.C.; Lee, P.J.; Bui, T.A.; Chang, B.H.; Lee, K.M. Superpixel algorithm for objects tracking in satellite video. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taiwan, China, 15–17 September 2021; pp. 1–2.
- 49. Zhang, Y.; Chen, D.; Zheng, Y. Satellite Video Tracking by Multi-Feature Correlation Filters with Motion Estimation. Remote Sens. 2022, 14, 2691.
- 50. Luo, W.; Xing, J.; Milan, A.; Zhang, X.; Liu, W.; Kim, T.K. Multiple object tracking: A literature review. Artif. Intell. 2021, 293, 103448.
- 51. Prokaj, J.; Medioni, G. Persistent tracking for wide area aerial surveillance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1186–1193.
- 52. Wei, J.; Sun, J.; Wu, Z.; Yang, J.; Wei, Z. Moving Object Tracking via 3D Total Variation in Remote-Sensing Videos. IEEE Geosci. Remote Sens. Lett. 2021, 19, 3506405.
- Zhang, J.; Jia, X.; Hu, J.; Tan, K. Satellite multi-vehicle tracking under inconsistent detection conditions by bilevel kshortest paths optimization. In Proceedings of the 2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 10–13 December 2018; pp. 1–8.
- 54. Ahmadi, S.A.; Ghorbanian, A.; Mohammadzadeh, A. Moving vehicle detection, tracking and traffic parameter estimation from a satellite video: A perspective on a smarter city. Int. J. Remote Sens. 2019, 40, 8379–8394.
- 55. Zhang, J.; Zhang, X.; Tang, X.; Huang, Z.; Jiao, L. Vehicle Detection and Tracking in Remote Sensing Satellite Vidio based on Dynamic Association. In Proceedings of the 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 5–7 August 2019; pp. 1–4.

Retrieved from https://encyclopedia.pub/entry/history/show/64334